ARTICLE **Communicated by Florian Luisier**

Bayes-Optimal Chemotaxis

Duncan Mortimer
dmorti@gmail.com
Queensland Brain Institute, University of Queensland, St. Lucia QLD 4072, Australia

Peter Dayan
dayan@gatsby.ucl.ac.uk
Gatsby Computational Neuroscience Laboratory, UCL, London WC1N 3AR, U.K.

Kevin Burrage
kevin.burrage@comlab.ox.ac.uk
Institute for Molecular Bioscience, University of Queensland, St. Lucia QLD 4072, Australia, and Computing Laboratory, Oxford University, Oxford OX1 3QD, U.K.

Geoffrey J. Goodhill
g.goodhill@uq.edu.au
Queensland Brain Institute and School of Mathematics and Physics, University of Queensland, St. Lucia QLD 4072, Australia

Chemotaxis plays a crucial role in many biological processes, including nervous system development. However, fundamental physical constraints limit the ability of a small sensing device such as a cell or growth cone to detect an external chemical gradient. One of these is the stochastic nature of receptor binding, leading to a constantly fluctuating binding pattern across the cell's array of receptors. This is analogous to the uncertainty in sensory information often encountered by the brain at the systems level. Here we derive analytically the Bayes-optimal strategy for combining information from a spatial array of receptors in both one and two dimensions to determine gradient direction. We also show how information from more than one receptor species can be optimally integrated, derive the gradient shapes that are optimal for guiding cells or growth cones over the longest possible distances, and illustrate that polarized cell behavior might arise as an adaptation to slowly varying environments. Together our results provide closed-form predictions for variations in chemotactic performance over a wide range of gradient conditions.

1 Introduction

Motion directed by an external inhomogeneous concentration field—chemotaxis—is an extremely common strategy used by many biological systems. For example, chemotactic behaviors are displayed by the slime mold *Dictyostelium discoideum* as it hunts for food and responds to stress from its environment, and neutrophils as they migrate to sites of inflammation (Downey, 1994; Parent & Devreotes, 1999; van Haastert & Devreotes, 2004). Chemotaxis also plays a crucial role in the development of the nervous system, allowing both migrating cells and axonal growth cones to find appropriate target regions (Song & Poo, 2001; Mortimer, Fothergill, Pujic, Richards, & Goodhill, 2008; Lowery & van Vactor, 2009). The importance of cellular chemotaxis for the correct functioning of organisms across many different contexts suggests that cells may have evolved sophisticated strategies for extracting as much information as possible from their environment about the direction of chemical gradients. However, their performance is ultimately limited by physical constraints, in particular noise intrinsic to the process of measuring an external gradient. The three most important sources of such noise are fluctuations in the external concentration as individual molecules diffuse, the stochastic nature of receptor-ligand binding, and fluctuations in signaling events inside the cell due to limited protein copy numbers (Berg & Purcell, 1977; Bialek & Setayeshgar, 2005).

Since the seminal work of Berg and Purcell (1977) investigating noise constraints on the measurement of local concentration, there has been much theoretical interest in understanding these fundamental limits to chemotaxis and gradient sensing (DeLisi & Marchetti, 1982; Lauffenburger, 1982; Tranquillo & Lauffenburger, 1986, 1987; Strong, Freedman, Bialek, & Koberle, 1998; Goodhill & Urbach, 1999; Bialek & Setayeshgar, 2005; Andrews & Iglesias, 2007; Ueda & Shibata, 2007; Kimmel, Salter, & Thomas, 2007; Endres & Wingreen, 2008, 2009; Fuller et al., 2010). Further modeling work has focused on potential biochemical implementations of chemosensory behavior in cells and growth cones (Hentschel & van Ooyen, 1999; Aeschlimann & Tettoni, 2001; Levchenko & Iglesias, 2002; Levine, Kessler, & Rappel, 2004; Sakumura, Tsukada, Yamamoto, & Ishii, 2005; Skupsky, Losert, & Nossal, 2005; Narang, 2006; Skupsky, McCann, Nossal, & Losert, 2007; Onsum & Rao, 2007; Naoki, Sakumura, & Ishii, 2008). However, none of this work has addressed the question of the strategy by which a gradient sensor should combine information from a spatially distributed array of receptors to optimally determine gradient direction. We took a step in this direction in Mortimer, Feldner, et al. (2009), inspired by the analogy with recent work in systems neuroscience examining Bayes-optimal strategies for making decisions in the face of uncertain sensory information (Green & Swets, 1966; Ernst & Banks, 2002; Knill & Pouget, 2004; Kording & Wolpert, 2004; Kording, 2007; Doya, Ishii, Pouget, & Rao, 2007). In particular, we determined analytically the optimal strategy for a one-dimensional gradient sensor to

determine whether its local concentration gradient increases toward the left or the right by formulating this as a problem in statistical decision theory. In addition, we performed a comprehensive experimental characterization of the response of axons to gradients and showed that variations in chemotactic performance as gradient steepness and concentration were systematically varied were consistent with the predictions of the Bayesian model.

In this article, we initially recapitulate the one-dimensional model, filling in many technical details that for reasons of space were omitted in Mortimer, Feldner, et al. (2009). We then significantly extend this theory in several ways. First, we show how these results predict a limit on how far a cell or growth cone can be guided by a single gradient. Second, we extend the model to predict how a cell or growth cone would respond optimally to multiple rather than just single gradients. This is an issue of key importance in nervous system development, where the correct guidance of axons depends on information gleaned from many different guidance cues. Third, we extend our model to two dimensions, a more realistic representation of receptor distributions, which allows a discussion of the model in the context of polarized versus unpolarized cells. Besides making several experimentally testable predictions, these results suggest more generally that a Bayesian approach to predicting optimal performance can be fruitful at the cellular as well as systems scale.

2 Theory

2.1 Gradient Sensing in One Dimension. We first describe the theory for a one-dimensional array of receptors, which establishes a framework for the extensions described later. We model individual cells as collections of chemoreceptors at positions $\{x, \ldots\} \subset \mathbf{R}$ and describe an ensemble of cells by giving a distribution over receptor arrangements $P_r(\{x\})$. We assume that the probability of a receptor being bound is determined by the equilibrium condition for Michaelis-Menten kinetics (Lauffenburger & Linderman, 1996) and that all receptors have identical dissociation constants K_d. Thus, for a receptor exposed to a local ligand concentration c, the probability that it is bound at any instant is

$$P(\text{bound}) = \frac{c/K_d}{1 + c/K_d}. \tag{2.1}$$

We further assume that the cell's local environment is completely described by a function $c(x)$, giving the concentration of ligand at position x. We focus on gradients that are sufficiently shallow that, on spatial scales on the order of a cell's diameter, they are effectively linear (i.e., second- and higher-order terms in the Taylor expansion of $c(x)$ are strongly dominated by the first-order term) and can be accurately described by a function of

the form $c(x) = c_0(1 + mx)$, where c_0 is the concentration at $x = 0$ and m is the relative rate of change of concentration with position—$m = \frac{1}{c_0}\frac{d}{dx}c(x)$. Equivalently, we use dimensionless units obtained by scaling concentration by K_d, defining $\gamma = c_0/K_d$ as the nondimensional concentration, and scaling distance by the diameter of the cell, l. Positions given relative to this length scale are denoted $r = x/l$, and the dimensionless gradient is $\mu = ml$. Thus, the probability that a receptor at position r is bound is

$$P(\text{bound}|\gamma, \mu, r) = \frac{\gamma \times (1 + \mu r)}{1 + \gamma \times (1 + \mu r)} \tag{2.2}$$

$$= \frac{\gamma}{1 + \gamma}\frac{1 + \mu r}{1 + \rho\mu r}, \tag{2.3}$$

and the probability that it is unbound is

$$P(\text{unbound}|\gamma, \mu, r) = 1 - P(\text{bound}|\gamma, \mu, r) \tag{2.4}$$

$$= \frac{1}{1 + \gamma}\frac{1}{1 + \rho\mu r}, \tag{2.5}$$

where ρ is the probability that a receptor at $r = 0$ is bound.

The states of the cell's receptors at any moment are described by two sets $\mathcal{B}$ and $\mathcal{U}$ containing the positions of the bound and unbound receptors, respectively. We assume that receptor binding occurs independently for different receptors, in which case the probability of obtaining such a state is

$$P(\mathcal{B}, \mathcal{U}|\gamma, \mu, \mathcal{R}) = C_{\mathcal{R}=\mathcal{B}\cup\mathcal{U}}(\mathcal{R})\left(\frac{\gamma}{1 + \gamma}\right)^n\left(\frac{1}{1 + \gamma}\right)^{N-n} \times \prod_{r\in\mathcal{B}}\frac{1 + \mu r}{1 + \rho\mu r}\prod_{r\in\mathcal{U}}\frac{1}{1 + \rho\mu r}, \tag{2.6}$$

where $n = |\mathcal{B}|$ is the number of bound receptors, $N = |\mathcal{R}|$ is the total number of receptors, and $C_{\mathcal{R}=\mathcal{B}\cup\mathcal{U}}(\mathcal{R})$ is 1 if $\mathcal{R} = \mathcal{B} \cup \mathcal{U}$ and 0 otherwise.

We assume that the cell does not have a priori access to complete information about the states and positions of its receptors when estimating the direction of the gradient. It seems unlikely that the cell will have an accurate record of where each of its possibly thousands of receptors are (though perhaps it might keep track of some summary statistic of receptor positons). Alternatively, the cell might precisely position its receptors (i.e., perhaps in a crystalline array). However, cell surface receptors have been observed to diffuse relatively freely (Ueda, Sako, Tanaka, Devreotes, & Yanagida, 2001; Tani et al., 2005). Thus, we assume that the cell must infer state and locations of its receptors on the basis of downstream signals that they produce. We treat in detail the case when only bound receptors produce signals, and these signals allow the cell to perfectly determine the positions of the signaling receptors. However, we also provide analogous results for the case

when only unbound receptors signal, and when both bound and unbound receptors produce distinguishable signals. We distinguish between these situations with this shorthand: $\mathcal{B}$ for only bound receptors signaling, $\mathcal{U}$ for only unbound receptors signaling, and $(\mathcal{B}, \mathcal{U})$ for both bound and unbound receptors signaling.

We model the cell's knowledge of its environment and the positions of its receptors (bound or unbound) with the prior distribution $P(\mu, \gamma, \mathcal{R})$. Bayes theorem then states that the cell's knowledge of the environment and its own configuration after observing receptor signals $\mathcal{B}$ is given by

$$P(\mu, \gamma, \mathcal{R}|\mathcal{B}) \propto P(\mathcal{B}|\mu, \gamma, \mathcal{R})P(\mu, \gamma, \mathcal{R}), \tag{2.7}$$

where $P(\mathcal{B}|\mu, \gamma, \mathcal{R}) = P(\mathcal{B}, \mathcal{R}/\mathcal{B}|\mu, \gamma, \mathcal{R})$. In other words, observing the signal $\mathcal{B}$ allows the cell to update the probability distribution representing its knowledge of the environment and its sensory state from $P(\mu, \gamma, \mathcal{R})$ to $P(\mu, \gamma, \mathcal{R}|\mathcal{B})$.

From this posterior distribution, the cell can make decisions about the gradient direction. A strategy for gradient sensing is then a function that relates available sensory input, $\mathcal{B}$, to a direction estimate $\delta = \pm 1$. Following the standard decision theory nomenclature, we refer to such mappings as policies. The aim of our analysis is to find the policy $\pi : \{\mathcal{B}\} \to \{\delta\}$, which is optimal in the sense that it maximizes a utility function $G(\delta; W)$. This utility function measures the usefulness of selecting a direction δ, given that the system is in a particular state, W. In our case, W corresponds to a particular instantiation of $\mathcal{R}$, γ, and μ. While there are many expressions that we might choose for $G(\delta; W) \equiv G(\delta; \mathcal{R}, \gamma, \mu)$, for simplicity we take

$$G(\delta; \mu, \gamma, \mathcal{R})_{attr} = \delta \times \text{sign}(\mu), \tag{2.8}$$

that is, the cell maximizes its utility by selecting the direction in which the gradient points, thus maximizing the correlation between δ and μ.

If the cell knew the state of its environment with certainty, it would be simple (in principle) for it to select its optimal action by choosing the action that maximizes the utility function. However, because the cell does not have perfect knowledge of the state of its environment, it must reason on the basis of the posterior distribution $P(\mu, \gamma, \mathcal{R}|\mathcal{B})$, given its sensory input $\mathcal{B}$, and prior information $P(\mu, \gamma, \mathcal{R})$. The Bayes optimal policy, $\pi_{bayes}(\mathcal{B})$ is defined as the policy that maximizes the expected value $E[G(\delta; \mu, \gamma, \mathcal{R})]$ of the utility function over the posterior distribution:

$$\pi_{bayes}(\mathcal{B}) = \underset{\delta}{\text{argmax}}\ E[G(\delta; \mu, \gamma, \mathcal{R})]_{P(\mu,\gamma,\mathcal{R}|\mathcal{B})} \tag{2.9}$$

$$= \underset{\delta}{\text{argmax}} \int_0^\infty d\gamma \int_{-\infty}^\infty d\mu \int d^N\mathcal{R} G(\delta; \mu, \gamma, \mathcal{R}) P(\mu, \gamma, \mathcal{R}|\mathcal{B}). \tag{2.10}$$

Since $G(\delta; \mu, \gamma, \mathcal{R})$ depends only on μ and δ, this can be simplified to

$$\pi_{bayes}(\mathcal{B}) = \underset{\delta}{\operatorname{argmax}}\ E[G(\delta; \mu)]_{P(\mu|\mathcal{B})}\,, \tag{2.11}$$

where

$$P(\mu|\mathcal{B}) = \int_0^\infty d\gamma \int d^N\mathcal{R}\, P(\mu, \gamma, \mathcal{R}|\mathcal{B}). \tag{2.12}$$

We approximate this strategy for the case when

1. The prior distribution factorizes, that is, $P(\mu, \gamma, \mathcal{R}) = P(\mu)P(\gamma)P(\mathcal{R})$,
2. The receptors are distributed independently of each other, that is, $P(\mathcal{R}) \propto \prod_{r\in\mathcal{R}} P(r)$—with $E[r] = 0$,
3. The cell expects a shallow gradient, that is, $P(\mu)$ is strongly concentrated near zero—in other words, $\int_{-\epsilon}^{\epsilon} d\mu\, P(\mu) \approx 1$ for some $\epsilon \ll 1$.
4. The cell has a sufficient number of receptors that it can estimate the background concentration essentially without error.

Under these assumptions we have

$$P(\mu|\mathcal{B}) \propto P(\mu)P(\mathcal{B}|\mu), \tag{2.13}$$

where

$$P(\mathcal{B}|\mu) \propto \int d^N\mathcal{R}\, P(\mathcal{R})P(\mathcal{B}|\mu, \gamma, \mathcal{R}) \tag{2.14}$$

$$\propto \int_{\mathcal{R}\supset\mathcal{B}} d^N\mathcal{R}\, P(\mathcal{R}) \prod_{r\in\mathcal{B}} \frac{1+\mu r}{1+\rho\mu r} \prod_{r\in\mathcal{R}/\mathcal{B}} \frac{1}{1+\rho\mu r} \tag{2.15}$$

$$= \left(\prod_{r\in\mathcal{B}} \frac{1+\mu r}{1+\rho\mu r}\right) \left(\int dr\, P(r) \frac{1}{1+\rho\mu r}\right)^{N-n}, \tag{2.16}$$

where we have split the integral over $\mathcal{R}$ into a product of integrals over individual receptors. Since $P(\mu)$ is concentrated near zero, we can approximate $P(\mathcal{B}|\mu)$ by taking a Taylor expansion of its logarithm about $\mu = 0$ to second order in μ:

$$\log P(\mathcal{B}|\mu) \approx \text{constant} + \mu \frac{\partial}{\partial\mu} \log P(\mathcal{B}|\mu)\Big|_{\mu=0} + \frac{\mu^2}{2} \frac{\partial^2}{\partial\mu^2} \log P(\mathcal{B}|\mu)\Big|_{\mu=0}. \tag{2.17}$$

Defining $R_b = \sum_{r\in\mathcal{B}} r$ and $S_b = \sum_{r\in\mathcal{B}} r^2$ and noting that $E[r] = 0$, we have

$$\frac{\partial}{\partial\mu}\log P(\mathcal{B}|\mu)\bigg|_{\mu=0} = \left(\sum_{r\in\mathcal{B}}\frac{r}{1+\mu r} - \rho\sum_{r\in\mathcal{B}}\frac{r}{1+\rho\mu r} - (N-n)\rho\int dr\, P(r)\frac{r}{(1+\rho\mu r)^2}\right)\bigg|_{\mu=0} \tag{2.18}$$

$$= (1-\rho)R_b, \tag{2.19}$$

and similarly,

$$\frac{\partial^2}{\partial\mu^2}\log P(\mathcal{B}|\mu)\Big|_{\mu=0} = -(1-\rho^2)S_b + 2(N-n)\rho^2 E\left[r^2\right]. \tag{2.20}$$

Thus, we obtain

$$P(\mu|\mathcal{B}) \propto P(\mu)\exp\left[(1-\rho)\,R_b\mu - \frac{1}{2}\left((1-\rho^2)S_b - 2(N-n)\rho^2 E\left[r^2\right]\right)\mu^2\right]. \tag{2.21}$$

In this expression, the first term in the exponential tends to shift the mean of the posterior distribution for μ in the direction of the center of mass of the bound receptors, R_b: larger values of R_b tend to indicate stronger gradients. As long as the term proportional to μ^2 is negative (which it almost surely is, provided ρ is not too close to 0 or 1 and N is sufficiently large), it contributes to reducing the variance of the posterior estimate of μ.

From this expression, we find the expected value of the utility function,

$$E[G(\delta;\mu)]_{P(\mu|\mathcal{B})} = \delta \times E\left[\text{sign}(\mu)\right]_{P(\mu|\mathcal{B})}. \tag{2.22}$$

Equation 2.22 obtains its maximum value when δ and $E\left[\text{sign}(\mu)\right]_{P(\mu|\mathcal{B})}$ have the same sign. If the cell is initially completely agnostic about the direction of the gradient, so that $P(\mu)$ is symmetric, then this decision is entirely determined by the sign of $R_\mathcal{B}$, so that $\pi_{bayes}(\mathcal{B}) = \text{sign}(R_\mathcal{B})$.

The performance of this strategy is given by $E\left[G(\pi_{bayes}(\mathcal{B});\mu)\right]_{P(\mathcal{B}|\mu,\gamma)} = \text{sign}(\mu)\times(P(R_\mathcal{B}>0|\gamma,\mu) - P(R_\mathcal{B}<0|\gamma,\mu))$. We will also refer to this quantity as $\Delta P(\gamma,\mu)$, the probability that μ and $R_\mathcal{B}$ have the same sign, minus the probability that they have opposite sign. Since $R_\mathcal{B}$ is the sum of many independent random variables, each of finite variance, the central limit theorem ensures that (for some range of γ and μ) $R_\mathcal{B}$ will be well approximated

by a gaussian distribution. We have

$$\begin{aligned} E[R_{\mathcal{B}}]_{P(\mathcal{B}|\mu,\gamma)} &= E\left[\sum_j r_j b_j\right] \\ &= NE\left[r\,E[b]_{P(b|\mu,\gamma,r)}\right]_{P(r)} \\ &= NE\left[r\rho + \mu\rho(1-\rho)r^2\right]_{P(r)} + O(\mu^2) \\ &= N\mu\rho(1-\rho)E\left[r^2\right]_{P(r)} + O(\mu^2), \end{aligned} \tag{2.23}$$

$$\begin{aligned} \mathrm{Var}(R_{\mathcal{B}})_{P(\mathcal{B}|\mu,\gamma)} &= N\mathrm{Var}(rb) \\ &= NE\left[r^2 E[b]_{P(b|\mu,\gamma,r)}\right]_{P(r)} - NE\left[r\,E[b]_{P(b|\mu,\gamma,r)}\right]^2_{P(r)} \\ &= \rho NE\left[r^2\right]_{P(r)} + O(\mu^2). \end{aligned} \tag{2.24}$$

We can then approximate the performance as

$$\begin{aligned} \Delta P(\mu,\gamma) &\approx \mathrm{sign}(\mu)\mathrm{erf}\left(\frac{E[R_{\mathcal{B}}]}{\sqrt{2\mathrm{Cov}(R_{\mathcal{B}}, R_{\mathcal{B}})}}\right) \\ &= |\mu|\sqrt{\frac{2}{\pi}}\sqrt{NE\left[r^2\right]_{P(r)}}\sqrt{\rho}(1-\rho) + O(\mu^2). \end{aligned} \tag{2.25}$$

We refer to the term

$$\mathrm{SNR}_{\mathcal{B}} = |\mu|\sqrt{\rho}(1-\rho) \tag{2.26}$$

as a signal-to-noise ratio (SNR). Expressed as a function of concentration, this is

$$\mathrm{SNR}_{\mathcal{B}} = |\mu|\sqrt{\frac{\gamma}{(1+\gamma)^3}}. \tag{2.27}$$

Thus, the model predicts a simple linear dependency of gradient sensing performance on gradient steepness μ, but a variation with concentration γ that is peaked at $\gamma = 1/2$ and falls off away from this value.

Up to an unknown constant of proportionality, this fits well to experimental data for growth cone chemotaxis to nerve growth factor as reported in Mortimer, Feldner et al. (2009). Furthermore, we obtain a good fit to previously published data from two other chemotactic systems: leukocyte chemotaxis to FMLLP (Zigmond, 1981) and *D. discoideum* chemotaxis to cyclic-AMP (Fisher, Merkl, & Gerisch, 1989); (see Figure 1). Finally, this model suggests that signals generated by peripheral receptors ought to be

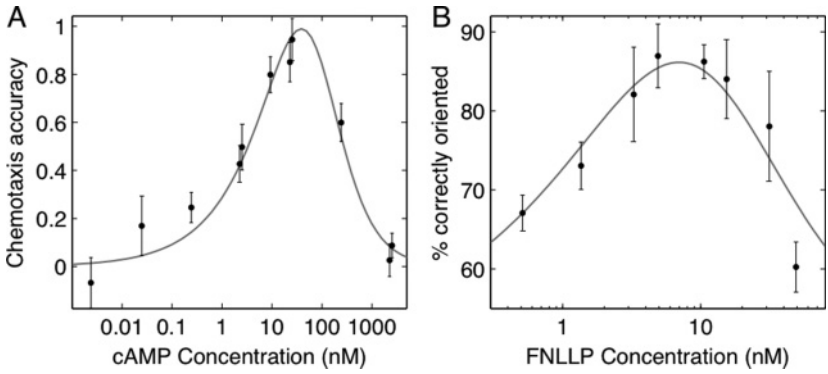

Figure 1: The gradient sensing model fits two different model systems in addition to axonal growth cones (Mortimer, Feldner, et al., 2009). (A) Chemotaxis accuracy of *D. discoideum* in response to cyclic AMP (data from Fisher et al., 1989). (B) Percentage of leukocytes orienting towards a gradient of n-Formylnorleucylleucylphenylalanine (FNLLP) (data from Zigmond, 1981). Experimental data are shown in black in each panel; error bars are 90% limits in A and SEMs in B. Gray curves are least-squares best fits of performance of the optimal strategy assuming only bound receptors produce localized signals. To fit, we assumed a linear relation between the experimental measures and the signal-to-noise ratio and allow the receptor dissociation constant K_d to vary. Best-fitting K_d values agree well with published experimental measurements.

weighted more highly than those generated by receptors located near the cell center (for a discussion of how this might be implemented biologically, see Mortimer, Feldner et al., 2009).

In this derivation, we focused on signals generated by bound receptors. However, in principle, unbound receptors might also produce signals in the absence of signaling from bound receptors or distinguishable from signals produced by bound receptors. This changes the information available to the cell and thus alters its gradient sensing performance. Following a similar process to that described above, it can be shown that when both bound and unbound receptors produce signals that can be distinguished from each other (so that in principle, the cell knows the locations of all of its receptors), gradient sensing performance is determined by a modified signal-to-noise ratio,

$$\mathrm{SNR}_{(\mathcal{B},\mathcal{U})} = |\mu|\sqrt{\rho(1-\rho)}, \tag{2.28}$$

while if only unbound receptors produce signals, performance is determined by

$$\mathrm{SNR}_{\mathcal{U}} = |\mu|\rho\sqrt{1-\rho}. \tag{2.29}$$

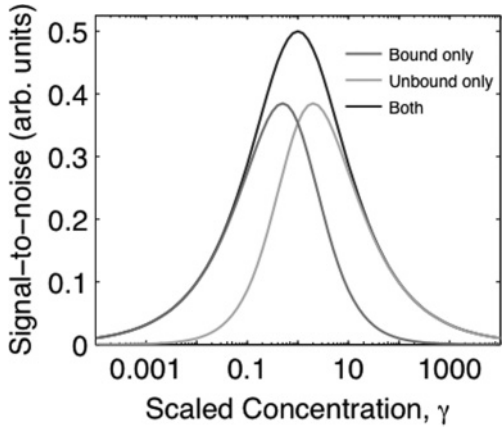

Figure 2: Gradient sensing performance depends on the signaling model. When only bound, or only unbound, receptors produce well-localized signals, performance is degraded at high or low concentrations, respectively, compared to the case when both bound and unbound receptors produce distinguishable signals.

Figure 2 illustrates the variation in performance with concentration for these three signaling conditions (equations 2.26, 2.28, and 2.29). As is clear from the figure, when both bound and unbound receptors produce signals, the gradient performance is better than when only bound or only unbound receptors produce signals. When only bound receptors produce signals, performance is degraded at high concentrations. This reduction in performance comes about because the cell is unable to distinguish between asymmetric signals due to the concentration gradient or signals stemming from asymmetries in how the receptors are distributed. These latter spurious signals play a larger role at high concentrations, thus reducing performance. This situation is reversed when only unbound receptors produce signals.

2.2 Maximum Distance for Guidance by a Single Gradient. We now use the theory above to predict the gradient shapes that would be optimal for guiding a chemotactic device over the longest possible distance (Goodhill & Baier, 1998). During neural development, chemical gradients are often involved in directing neurons and their axons to their appropriate target locations. In order to achieve robust developmental repeatability, presumably a threshold level of allowable error can be accepted: as long as errors are not made too frequently, they can be corrected by subsequent behavior. In general, we would expect there to be an inverse relationship between the maximum possible distance over which guidance can occur, and the level of error that can be tolerated, for a fixed investment of resources (such as number of receptors, or cell size) in the cell's sensory equipment.

While it seems unlikely that we will be able to determine how an organism might trade off between cell sensory performance and energy that might otherwise be used elsewhere, we can consider what general features we would expect to see in gradients that are designed to provide guidance over the largest possible distance to within a tolerable level of error to cells with fixed sensory capabilities.

In previous work we considered what gradient shapes would be optimal in this sense under simpler models of the constraints underlying chemotaxis (Goodhill, 1997; Goodhill & Urbach, 1999). Under the Bayesian framework, the idea of a maximum error rate or threshold for successful guidance naturally maps to a threshold on the signal-to-noise ratio. Thus, we consider a gradient $\gamma(x)$ to provide effective guidance over a region Ω, when, for each position x within Ω, the signal-to-noise ratio is greater than or equal to some threshold θ. In other words, no matter where the cell is within the gradient, the gradient provides enough information to the cell for its error rate to remain below threshold.

Substituting $\mu = \frac{l}{\gamma}\frac{d\gamma}{dx} = l\frac{1}{\rho(1-\rho)}\frac{d\rho}{dx}$ into equation 2.25 and comparing with the threshold θ, we find that effective guidance can be achieved whenever

$$\frac{l}{\epsilon}\frac{1}{\sqrt{\rho}}\frac{d\rho}{dx} \geq \theta, \tag{2.30}$$

where l is the width of the cell (and thus sets the relevant length scale for the problem) and $\epsilon = \sqrt{\pi/2NE\left[r^2\right]_{P(r)}}$.

The gradient that would provide guidance over the longest distance is therefore obtained by enforcing equality in equation 2.30. Integrating the resulting differential equation yields an expression for the one-dimensional distribution of guidance cues that achieves this theoretically optimal performance:

$$\int_0^{\rho} d\rho' \frac{1}{\sqrt{\rho'}} = 2\sqrt{\rho} = \frac{x}{l/\epsilon\theta}. \tag{2.31}$$

Rearranging to get concentration as a function of distance, we have

$$\gamma_{\mathcal{B}}(x) = \frac{\left(\frac{x}{2l/\epsilon\theta}\right)^2}{1 - \left(\frac{x}{2l/\epsilon\theta}\right)^2}. \tag{2.32}$$

On inspection (and illustrated in Figure 3), this expression asymptotes to infinity as x approaches $X = \frac{2l}{\epsilon\theta}$. In other words, the concentration of the guidance cue must rise more and more rapidly as x approaches X, so that in fact, it is impossible to exceed the distance X. Thus, even neglecting practical constraints on the maximum concentration of guidance cue that

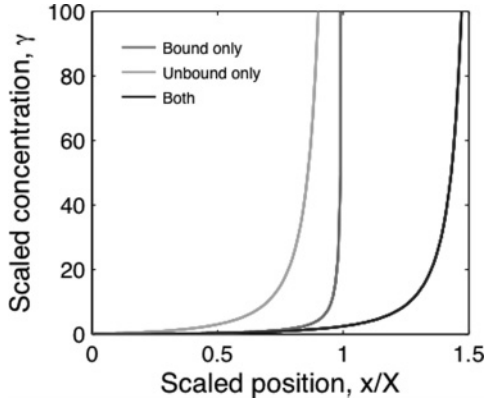

Figure 3: Optimal gradients for guidance over long distances given the three different signaling mechanisms. Both $\mathcal{B}$ (optimal gradient shape shown in dark gray) and $\mathcal{U}$ (in light gray) can guide the cell over the same maximum distance X, represented on the figure as $x/X = 1$. The optimal gradient for guidance by the $(\mathcal{U}, \mathcal{B})$ strategy is shown in black.

can be produced, X is the maximum distance that can be obtained given the $\mathcal{B}$ signaling mechanism, with a threshold for effective guidance of θ.

Similar arguments yield optimal gradients for the $\mathcal{U}$ and $(\mathcal{B}, \mathcal{U})$ signaling mechanisms: making use of equations 2.29 and 2.28, we can write expressions analogous to equation 2.30 for the $\mathcal{U}$ and $(\mathcal{U}, \mathcal{B})$ cases: For $\mathcal{U}$,

$$\frac{l}{\epsilon}\frac{1}{\sqrt{1-\rho}}\frac{d\rho}{dx} \geq \theta, \tag{2.33}$$

and for $(\mathcal{U}, \mathcal{B})$,

$$\frac{l}{\epsilon}\frac{1}{\sqrt{\rho(1-\rho)}}\frac{d\rho}{dx} \geq \theta. \tag{2.34}$$

Again, these differential equations can be integrated by separation of variables and rearranged to express the optimal gradient in terms of concentration, yielding for $\mathcal{U}$,

$$\gamma_{\mathcal{U}}(x) = \frac{1}{(1-x/X)^2} - 1, \tag{2.35}$$

and for $(\mathcal{U}, \mathcal{B})$,

$$\gamma_{\mathcal{U},\mathcal{B}}(x) = \tan^2(x/X). \tag{2.36}$$

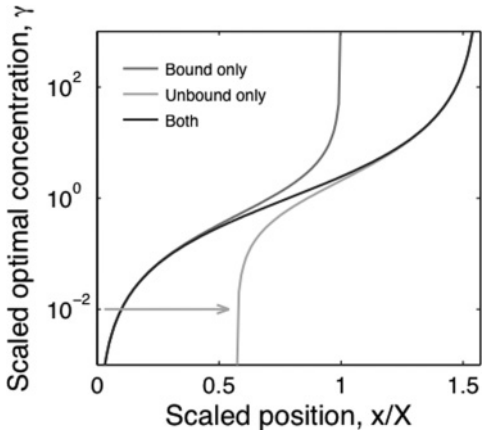

Figure 4: Symmetry in the optimal gradient shapes for the three signaling mechanisms. The data in Figure 3 are replotted on a log scale, with the curve for $S = \mathcal{U}$ shifted to the right by $\pi/2 - 1$ (as indicated by the arrow) in order to demonstrate that it converges to the optimal gradient for $S = (\mathcal{B}, \mathcal{U})$ at high concentrations.

The optimal gradient shapes for the three signaling mechanisms ($\mathcal{B}$, $\mathcal{U}$, and $(\mathcal{U}, \mathcal{B})$) given by equations 2.32, 2.35, and 2.36 are plotted in Figure 3. In both the $\mathcal{B}$ and $\mathcal{U}$ cases, the concentration asymptotes to infinity at the same value, X, representing the maximum possible guidance distance by these gradients. It is unsurprising that these two mechanisms both provide the same maximum guidance distance owing to the symmetry of the two situations—for example, the optimal strategy and performance for the $\mathcal{B}$ signaling mechanism can be mapped to that for the $\mathcal{U}$ mechanism by everywhere exchanging b for $1 - b$ (similarly, ρ for $1 - \rho$). For the $(\mathcal{B}, \mathcal{U})$ case, guidance can be achieved over a longer distance, $\frac{\pi}{2} X$, consistent with the fact that more information is available to the cell in this situation. Changing the threshold θ does not change the shape of the optimal gradient, but rather rescales the gradient by stretching or compressing it along the x-axis. Similarly, the maximum achievable guidance distance scales with $\sqrt{N}$, the square root of the number of chemoreceptors, though again the overall shape of the gradient is unaffected.

The symmetry of the $\mathcal{B}$ and $\mathcal{U}$ situations is illustrated in Figure 4: this demonstrates that the optimal gradient shapes for $\mathcal{B}$ and $\mathcal{U}$ attain the same shape as the $(\mathcal{B}, \mathcal{U})$ gradient in the low-concentration and high-concentration limits, respectively.

Note that for the $\mathcal{B}$ case, a much lower maximum ligand concentration is required to reach any given distance, as compared to the $\mathcal{U}$ case, as is illustrated in Figure 5. The $\mathcal{B}$ signaling mechanism is therefore more

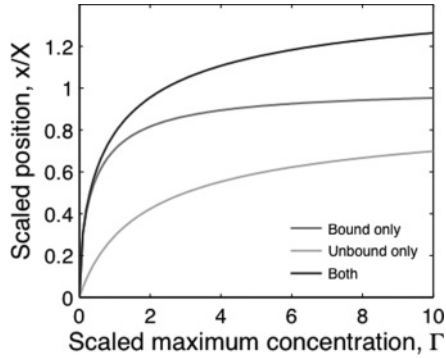

Figure 5: Optimal gradients for $\mathcal{B}$ and $\mathcal{U}$ approximate the $(\mathcal{U}, \mathcal{B})$ gradient at low and high concentrations, respectively. Here, the maximum guidance distance possible is plotted given a maximum allowed concentration, γ. Distances are expressed as fractions of the maximum possible guidance distance X for the $S = \mathcal{B}$ and $S = \mathcal{U}$ cases. As can be seen in this figure, $S = \mathcal{B}$ approaches its saturating value much more rapidly than $S = \mathcal{U}$, suggesting that effective long-range guidance can be achieved without requiring the expression of as much ligand as is required for the $S = \mathcal{U}$ case.

"efficient" for guidance over long range, in that less guidance cue is required to achieve a given distance. This is because when only bound receptors signal, the cell is more sensitive at low than high concentrations, and hence a greater proportion of the maximum guidance distance is achieved at low concentrations. This is reversed for the $\mathcal{U}$ case.

This improved efficiency suggests that localized signaling by bound receptors may be preferred over signaling from unbound receptors for chemotaxing structures such as growth cones and neutrophils, which follow gradients established in vivo, as less guidance cue is required. Axon guidance molecules are typically large, and hence metabolically expensive proteins. By making use of a $\mathcal{B}$ signaling strategy for its growth cones, an organism may achieve guidance over a given distance for a lower metabolic cost.

The $(\mathcal{U}, \mathcal{B})$ strategy does better than both of the other signaling strategies, as we would expect given that more information is available to the cell in this case. However, as is apparent from Figure 5, much of the additional effective guidance distance provided in principle by the $(\mathcal{U}, \mathcal{B})$ signaling strategy over the $\mathcal{B}$ strategy requires a great increase in concentration over that required to achieve within a few percent of the maximum guidance distance achieved by the $\mathcal{B}$ strategy. Coupled with additional complexity in the signaling mechanism and intracellular machinery that would presumably be required to implement the $(\mathcal{B}, \mathcal{U})$ strategy, this strategy may become less desirable despite its greater in-principle maximum guidance distance.

A key assumption made in deriving equations 2.32, 2.35, and 2.36 is that the cell or growth cone does not adaptively modify its sensory apparatus as the gradient conditions change in order to improve its performance. For example, it could do this by expressing more receptors on its surface, modifying the kinetics of its receptors, or increasing its spatial extent. In this case, we would expect that rather than observing a single optimal gradient shape, we would observe correlations between local gradient conditions and the properties of the cell. Another way in which the results obtained above might be "outperformed" would be if the cell or growth cone were able to tolerate occasional sharp decreases in concentration while maintaining some memory of the correct growth direction. How this would be achieved would depend on details of the cell or growth cone's motion. For example, memory of past gradient estimates might be maintained by a polarized intracellular molecule with slow dynamics (Arrieumerlou & Meyer, 2005), or the cell might actively ignore concentration changes that occur too rapidly. This would allow the gradient to be "reset," in principle allowing for guidance over unlimited distance.

2.3 Optimal Responses to Multiple Gradients. So far we have considered only guidance by a single gradient. However, cells or growth cones may be exposed to multiple chemotactic cues simultaneously (Song & Poo, 2001; Hao et al., 2001; Heit, Tavener, Raharjo, & Kubes, 2002; Charron, Stein, Jeong, McMahon, & Tessier-Lavigne, 2003; McLaughlin & O'Leary, 2005). In this case, they must somehow combine the information obtained from each gradient. To see where such a combination of cues might be useful, we consider how far a growth cone might be guided by a combination of two guidance cues. It is easy to see that two cues offer immediate advantages over a single cue. Consider two growth cones, both with a total of N guidance cue receptors. However, the first has receptors for only one kind of guidance cue molecule, while the second has receptors for two different kinds of molecule, each of which is perfectly distinguishable by the receptors (i.e., there is little or no cross-reactivity). Then, denoting the maximum distance achievable by the first growth cone (i.e. guidance by a single guidance cue) by X, we can ask whether we might achieve guidance over a longer distance by using multiple guidance cues. While we do not know the optimal arrangement of guidance cue concentrations for achieving this task, it is easy to see that guidance over a longer distance can be achieved by simply placing two optimal gradients for each guidance cue by itself, one after the other. Since the total number of receptors is N, we know that the number of receptors for the first type of guidance cue, and that for the second, must sum to N. Since the maximum guidance distance scales with the square root of the number of receptors, we have

$$X_{N_1,N_2} = \sqrt{\frac{N_1}{N}}X + \sqrt{\frac{N_2}{N}}X = \left(\sqrt{N_1/N} + \sqrt{1 - N_1/N}\right)X. \tag{2.37}$$

This is maximized when $N_1 = N_2 = N/2$, achieving

$$X_{N/2,N/2} = \sqrt{2}X > X. \tag{2.38}$$

In other words, by stacking two optimal gradients for different cues one after the other and assuming an equal share of receptors for the two classes of guidance cue, we can increase the maximum guidance distance by at least a factor of $\sqrt{2}$.

We consider the simplest possible extension of our one-cue model to multiple cues: we assume that there are now two classes of receptor, each perfectly selective for just one of two guidance cues. We also assume that only bound receptors produce well-localized signals. Treating the receptors independently and identifying the two guidance cues and their receptors by subscripts 1 and 2, we can extend the likelihood function to include both cues:

$$\begin{aligned} &P(\mathcal{B}_1, \mathcal{B}_2|\gamma_1, \mu_1, \gamma_2, \mu_2, \mathcal{R}_1, \mathcal{R}_2) \\ &\quad = P(\mathcal{B}_1|\gamma_1, \mu_1, \mathcal{R}_1)P(\mathcal{B}_2|\gamma_2, \mu_2, \mathcal{R}_2) \qquad (2.39) \\ &\quad = \frac{\gamma_1^{n_1}}{(1+\gamma_1)^{N_1}} \prod_{r\in\mathcal{B}_1} \frac{1+\mu_1 r}{1+\rho_1\mu_1 r} \prod_{r\in\mathcal{R}_1/\mathcal{B}_1} \frac{1}{1+\rho_1\mu_1 r} \\ &\qquad \times \frac{\gamma_2^{n_2}}{(1+\gamma_2)^{N_2}} \prod_{r\in\mathcal{B}_2} \frac{1+\mu_2 r}{1+\rho_2\mu_2 r} \prod_{r\in\mathcal{R}_2/\mathcal{B}_2} \frac{1}{1+\rho_1\mu_1 r}, \qquad (2.40) \end{aligned}$$

where $n_i = |\mathcal{B}_i|$ and we have used dimensionless concentrations $\gamma_i = C_i/K_{di}$ and $\rho_i = \gamma_i/(1+\gamma_i)$. $\mathcal{B}_i$ and $\mathcal{R}_i$ are multisets, containing the positions of only bound receptors, and both bound and unbound receptors, respectively.

As is apparent from equation 2.40, we are making a straightforward extension of the one-gradient case in order to describe the local distribution of the two guidance molecules. For each cue, we have assumed that on the length scale relevant to a single cell, the local spatial distribution of cue can be approximated by

$$\gamma_i(r) = \gamma_i(0)\,(1+\mu_i r)\,, \tag{2.41}$$

where $\gamma_i(r)$ is the concentration of guidance cue i at relative position r, scaled by the dissocation constant of its cognate receptor (thus, $\gamma_i(0)$ is the concentration at the center of the cell), and μ_i is the relative change in γ_i across the width of the cell. Hence, in order to describe the cell's prior distribution for its environment, we need a probability distribution over four parameters: $P(\gamma_1, \mu_1, \gamma_2, \mu_2)$.

To constrain this distribution, we focus on situations in which the guidance cues are distributed in a distinct spatial pattern, such that they all vary along the same direction, each either increasing or decreasing in concentration. In other words, the gradients are correlated with each other. Since the cell has coevolved with this arrangement of gradients, it can make use of these correlations to improve its estimates of the direction of each gradient individually. If the cues were entirely independent of one another (and the cell "knew" this), the prior would be of the form $P(\gamma_1, \mu_1)P(\gamma_2, \mu_2)$. On the other hand, for the situation we are particularly interested in, a simple expression for $P(\gamma_1, \mu_1, \gamma_2, \mu_2)$ can be obtained by making as many independence assumptions as are reasonable while retaining structure in the prior that specifies that the two gradients always point in the same directions. This puts only one constraint on the prior; that the signs of the gradients are perfectly correlated. So, for example, if the gradients are known to always point in the same direction, we can write the prior in the form

$$P(\gamma_1, \mu_1, \gamma_2, \mu_2) = \begin{cases} P(\gamma_1, \gamma_2, |\mu_1|, |\mu_2|) & \text{if } \operatorname{sign}(\mu_1) = \operatorname{sign}(\mu_2), \textit{or} \\ 0 & \text{if } \operatorname{sign}(\mu_1) \neq \operatorname{sign}(\mu_2). \end{cases} \tag{2.42}$$

In this case, the prior takes a positive value only if both μ_1 and μ_2 have the same sign (in which case the prior depends on only the magnitudes of the two gradients)—that is point in the same direction. Similarly, for two gradients known to point in opposite directions, an appropriate prior would be

$$P(\gamma_1, \mu_1, \gamma_2, \mu_2) = \begin{cases} P(\gamma_1, \gamma_2, |\mu_1|, -|\mu_2|) & \text{if } \operatorname{sign}(\mu_1) = -\operatorname{sign}(\mu_2), \text{ or} \\ 0 & \text{otherwise.} \end{cases} \tag{2.43}$$

Applying Bayes' theorem and taking a Taylor expansion to first order in μ_1 and μ_2, we obtain an approximate expression for the posterior distribution of μ_1 and μ_2, given $\mathcal{B}_1$ and $\mathcal{B}_2$:

$$P(\mu_1, \mu_2|\mathcal{B}_1, \mathcal{B}_2) \propto P(\mu_1, \mu_2)\,(1 + \mu_1(1 - \hat{\rho}_1)R_{b1} + \mu_2(1 - \hat{\rho}_2)R_{b2}) \tag{2.44}$$

where we have assumed that $P(\mathcal{R}_1, \mathcal{R}_2) = P(\mathcal{R}_1)P(\mathcal{R}_2)$, and $P(\mu_1, \mu_2, \gamma_1, \gamma_2) = P(\mu_1, \mu_2)P(\gamma_1, \gamma_2)$ and that $P(\mu_1, \mu_2)$ is nonzero only for small μ_1 and μ_2. Here, $\hat{\rho}_i$ represents the best *a posteriori* estimate for ρ_i, assuming $\mu_1 = \mu_2 = 0$. Note that in general, under these assumptions, $\hat{\rho}_1$ and $\hat{\rho}_2$ depend on the number of bound molecules of both types of ligand; this is because γ_1 and γ_2 would in general be correlated, and hence

obtaining information about γ_1 also gives information about γ_2. Under the additional assumption that these concentrations are independent—that is, the prior for the concentrations factors, $P(\gamma_1, \gamma_2) = P(\gamma_1)P(\gamma_2)$ — $\hat{\rho}_1$ and $\hat{\rho}_2$, are dependent on only the binding of the related ligand.

We now make the key assumption that, as described earlier, the directions of the gradients are not independent—in other words, that

$$P(\mu_1, \mu_2) = \begin{cases} P(|\mu_1|, |\mu_2|) & \text{if } \operatorname{sign}(\mu_1) = \operatorname{sign}(\mu_2), \text{ or} \\ 0 & \text{if } \operatorname{sign}(\mu_1) \neq \operatorname{sign}(\mu_2) \end{cases}, \tag{2.45}$$

or, equivalently, $P(\mu_1, \mu_2) \equiv P(a)P(|\mu_1|)P(|\mu_2|)$ where $\mu_1 = a|\mu_1|$ and $\mu_2 = a|\mu_2|$. Thus, while the gradient strengths are independent, they always point in the same direction a.

The cell is thus trying to infer a from signals $\mathcal{B}_1$ and $\mathcal{B}_2$. We need to calculate $P(a|\mathcal{B}_1, \mathcal{B}_2)$. This is achieved by integrating equation 2.44 with respect to the magnitudes of the two gradient steepnesses:

$$\begin{aligned} P(a|\mathcal{B}_1, \mathcal{B}_2) &= \int_0^\infty d\mu_1 \int_0^\infty d\mu_2 P(a)P(|\mu_1|)P(|\mu_2|)P(a|\mu_1|, a|\mu_2| \big| \mathcal{B}_1, \mathcal{B}_2) \\ &\approx \int_0^\infty d\mu_1 \int_0^\infty d\mu_2 P(a)P(|\mu_1|)P(|\mu_2|)\,(1 + a|\mu_1|(1 - \hat{\rho}_1)R_{b1} \\ &\quad + a|\mu_2|(1 - \hat{\rho}_2)R_{b2}) \\ &= P(a)\left(1 + a \sum_{i=1..2} \langle|\mu_i|\rangle(1 - \hat{\rho}_i R_{bi})\right). \end{aligned} \tag{2.46}$$

Assuming that $a = \pm 1$ are equally likely a priori, the most likely a posteriori value for a is

$$a_{MAP} = \operatorname{sign}(\langle|\mu_1|\rangle(1 - \hat{\rho}_1)R_{b1} + \langle|\mu_2|\rangle(1 - \hat{\rho}_2)R_{b2}). \tag{2.47}$$

In other words, the cell's optimal decision about the appropriate direction to turn is made by comparing $\sum_{i=1..2}\langle|\mu_i|\rangle(1 - \hat{\rho}_i)R_{bi}$ with zero. This simply extends the single gradient case by linearly combining the contributions for the two cues, with prefactors that discount information at high concentrations. The contribution to the decision of each signal R_{bi} from its cognate receptors should be reduced according to the term $1 - \hat{\rho}_i$, where $\hat{\rho}_i$ is the cell's best estimate of the fraction of the corresponding receptors that are bound. This "preprocessing" improves performance because at high concentrations of cue i (when $(1 - \hat{\rho}_i) \approx 0$), the input R_{bi} is dominated by biases due to the inevitable fluctuations in receptor density.

It is also worth noting that the extension to more than two cues follows along identical lines. If M guidance cues are present, all pointing along the same axis, then the optimal decision is to compare

$$\chi = \sum_{i=1}^{M} \langle \mu_i \rangle (1 - \hat{\rho}_i) R_{bi}$$

with zero, where the subscript i refers to the ith such gradient.

We now calculate the performance of this two-gradient decision strategy, again, by estimating the probability of the cell making the correct choice of gradient direction. We assume that the cell expects two gradients to be presented, with both pointing along the same axis. With the additional assumption that the cell's prior $P(\gamma_1, \gamma_2, \mu_1, \mu_2)$ expresses no correlation between gradient steepness and concentration for the two gradients, the optimal decision is obtained by comparing $\chi = \langle |\mu_1| \rangle (1 - \hat{\rho_1}) R_{b1} + \langle |\mu_2| \rangle (1 - \hat{\rho_2}) R_{b2}$ with zero. As it is experimentally feasible to present opposed gradients, despite the cell's expectations, we consider the cell to have made the correct decision if it elected to turn in the direction of the first cue. Hence, we wish to obtain an analytic expression for $P_{correct}(\gamma_1, \gamma_2, \mu_1, \mu_2) = P(\chi > 0 \mid \gamma_1, \gamma_2, \mu_1 > 0, \mu_2) + P(\chi < 0 \mid \gamma_1, \gamma_2, \mu_1 < 0, \mu_2)$.

We assume that the prior for γ_1 and γ_2 satisfies $P(\gamma_1, \gamma_2) = P(\gamma_1)P(\gamma_2)$ and that the individual priors are beta distributions in ρ_i, that is,

$$P(\gamma_i)\, d\gamma_i = P(\rho_i) d\rho_i = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \rho_i^{\alpha_i - 1} (1 - \rho_i)^{\beta_i - 1} d\rho_i, \tag{2.48}$$

in which case $\hat{\rho}_i = \frac{n_i + \alpha_i}{N_i + \beta_i}$. Here, $\Gamma(x)$ is Euler's gamma function (Abramowitz & Stegun, 1970). We make this choice of prior distribution as it enables us to explicitly calculate a value for $\hat{\rho}_i$, which is not possible in general.

Again, for suitably large values of γ_1 and γ_2, we can approximate $P(\chi | \gamma_1, \gamma_2, \mu_1, \mu_2)$ with a normal distribution with mean $\langle \chi \rangle$ and variance $\langle \chi^2 \rangle - \langle \chi \rangle^2$. Following similar but more involved calculations than for the one-cue case in section 2.1, we arrive at the results:

$$\langle \chi \rangle \approx \sum_{i=1}^{2} \frac{1}{12} \left(\frac{N_i - 1 + \alpha_i}{N_i + \beta_i} - \frac{N_i - 1}{N_i + \beta_i} \rho_i \right) (1 - \rho_i) \rho_i N_i \mu_i \tag{2.49}$$

$$\approx \frac{1}{12} \left(N_1 (1 - \rho_1)^2 \rho_1 \mu_1 + N_2 (1 - \rho_2)^2 \rho_2 \mu_2 \right) \tag{2.50}$$

$$\langle \chi^2 \rangle - \langle \chi \rangle^2 \approx \sum_{i=1}^{2} \frac{1}{12} \left(\left(\frac{N_i - 1 + \alpha_i}{N_i + \beta_i} \right)^2 - 2 \frac{N_i - 1}{N_i + \beta_i} \frac{N_i - 2 + \alpha_i}{N_i + \beta_i} \rho_i + \right. \tag{2.51}$$

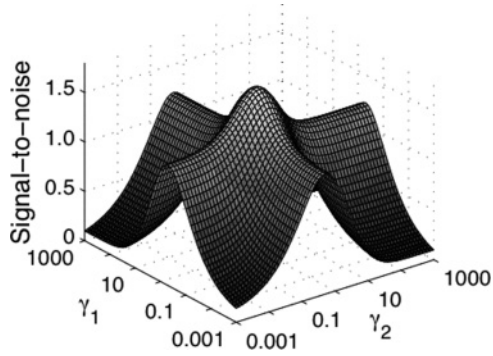

Figure 6: Sensing performance when presented with two gradients indicating the same target direction. The probability of the cell's making a correct decision about the gradient direction is plotted against the concentrations of two ligands.

$$\frac{N_i - 1}{N_i + \beta_i}\frac{N_i - 2}{N_i + \beta_i}\rho_i^2\Big)\,\rho_i N_i \tag{2.52}$$

$$\approx \frac{1}{12}\left(N_1(1-\rho_1)^2\rho_1 + N_2(1-\rho_2)^2\rho_2\right), \tag{2.53}$$

where we have expanded the expressions to first order in μ_i and assumed that $N_i \gg \alpha_i$, $N_i \gg \beta_i$ and $N_i \gg 1$.

Hence, we have that the probability of selecting the correct direction is given by

$$P_{correct}(\gamma_1, \gamma_2, \mu_1, \mu_2) \approx \frac{1}{2}\mathrm{erfc}\left(-\left|\frac{1}{\sqrt{24}}\frac{N_1(1-\rho_1)^2\rho_1\mu_1 + N_2(1-\rho_2)^2\rho_2\mu_2}{\sqrt{N_1(1-\rho_1)^2\rho_1 + N_2(1-\rho_2)^2\rho_2}}\right|\right), \tag{2.54}$$

where $\mathrm{erfc}(x)$ is the complementary error function. This combination rule leads to the sensitivity surface illustrated in Figure 6, which shows the fraction of times the correct turning decision is made under the influence of two gradients of the same steepness as a function of their background concentrations. This shape is consistent with the performance obtained by a cell exposed to only one guidance cue, as can be seen by considering the appropriate limits ($\gamma_1 = 0$ or $\gamma_2 = 0$).

Suppose the cell does not optimally weight the signals from its receptors, but instead makes its decision based on the unweighted sum of the signals: $R_{b1} + R_{b2}$. Noise would then reduce performance when the concentration of one or both cues is high. Experimentally, we could distinguish between this, and the optimal rule, by presenting a gradient of one cue over

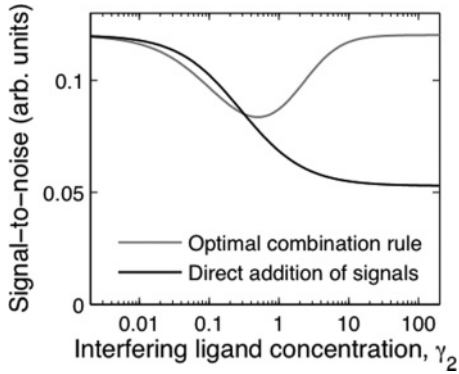

Figure 7: Predicted experimental sensitivity curve trends when a gradient of one guidance cue is presented over various uniform "interfering" backgrounds of a second cue. The concentration of the interfering cue is varied over the range indicated on the x-axis. The sensitivity curve for the optimal combination strategy (see equation 2.47) is in light gray; the black curve is predicted when signals are combined without accounting for noise. In each case, the number of receptors for the two cues is equal.

a uniform background concentration of the other (see Figure 7). Increasing the concentration of the second cue initially degrades performance for both strategies: since no gradient of this cue is present, noise is progressively added to the system. Counterintuitively, the suboptimal strategy "outperforms" the optimal strategy at low concentrations. This occurs because both strategies assume that both the informative and interfering guidance cue are providing useful information. At low concentrations of the interfering cue, the optimal strategy gives greater relative weight to the interfering cue than the informative cue, effectively increasing the noise in the response and reducing the growth cone's ability to sense the gradient. This occurs only because we are probing these strategies with an artificial situation not modeled by the prior distribution.

As the concentration increases further, however, the optimal combination rule regains sensitivity as the signals from the second cue are discounted. In contrast, if the cues are combined simply by adding signals, performance is further reduced in this case as more noise is introduced through R_{b2}. We might expect to observe the suboptimal direct addition combination rule when an organism is attempting to respond to cues not usually present together in an ecologically relevant setting—under such conditions, there would be no evolutionary pressure toward discounting one cue in favor of the other. In contrast, we may observe the optimal combination rule in systems where the two cues are usually present simultaneously.

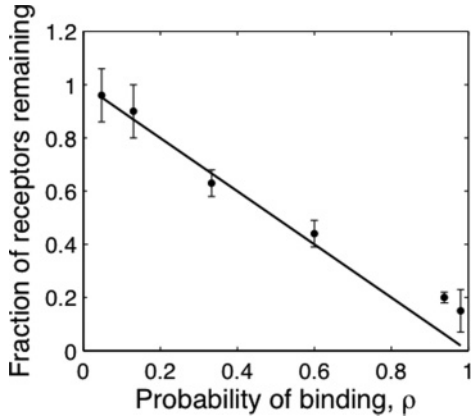

Figure 8: Leukocyte receptor downregulation is consistent with predictions of the optimal cue combination strategy. Leukocytes downregulate their receptors in response to stimulation with FNLLP. The points on the graph show the fraction of receptors that remains on the surface plotted against the fraction ρ of receptors that would be bound at a given FNLLP concentration (calculated from $[FNLLP]/([FNLLP] + K_d)$) (data from Zigmond, 1981). The solid line shows the prediction from our model, assuming that receptor downregulation is the primary method by which leukocytes discount signals from a class of receptors that are mostly bound. We obtain a good fit with no free parameters.

The preprocessing advocated by the optimal strategy could be implemented through an adaptation process in which the strength of signals generated by a receptor depends on the average ligand binding to that receptor. The ideal observer approach makes quantitative predictions concerning the nature of this adaptation process: specifically, that the contribution of a class of receptors to downstream signaling pathways should be proportional to $(1 - \rho)$. G-protein coupled receptors are known to display such adapting behavior: sustained signaling leads to receptor phosphorylation and, subsequently, reduced downstream activity (Kohout & Lefkowitz, 2003).

Alternatively, adaptation might be achieved by downregulating receptors at high concentrations. Indeed, this has been observed in leukocytes exposed to high concentrations of FNLLP (Zigmond, 1981). Furthermore, according to the optimal combination rule, receptors should be downregulated so that the fraction of receptors remaining on the surface of the growth cone is proportional to $(1 - \rho)$. Figure 8 shows that, remarkably, receptor downregulation data for polymorphonuclear neutrophils (Zigmond, 1981) give an excellent match to this prediction, where there are no free parameters. We suggest that adaptation or receptor downregulation may play a role in optimally combining noisy sources of information, regardless of subsequent downstream processing.

2.4 Gradient Sensing in Two Dimensions. So far we have considered only a one-dimensional array of receptors. While this is clearly a useful approximation and leads to predictions that fit well with experimental data (Mortimer, Feldner, et al., 2009), in reality cells and growth cones are two-dimensional. It is thus natural to extend the model to a two-dimensional distribution of the receptors and ask whether this changes the predictions of the model.

It is in fact straightforward to extend the results of section 2.1 to two dimensions. Receptor positions are now two-dimensional column vectors, $\boldsymbol{r}$, and the gradient steepness is also a column vector, $\boldsymbol{\mu} = \frac{l}{c_0}\nabla c(0)$, so that the concentration at position $\boldsymbol{r}$ is

$$\gamma(\boldsymbol{r}) = \gamma \times (1 + \boldsymbol{\mu} \cdot \boldsymbol{r}), \tag{2.55}$$

or, equivalently,

$$\gamma(\boldsymbol{r}) = \gamma \times \left(1 + \boldsymbol{r}^T \boldsymbol{\mu}\right). \tag{2.56}$$

Furthermore, as the cell can now elect to step in any direction, θ, rather than simply turn to the left or right, we redefine the utility function as

$$G(\theta; \boldsymbol{\mu}) = \boldsymbol{u}_\theta^T \boldsymbol{\mu}, \tag{2.57}$$

where $\boldsymbol{u}_\theta$ is the unit vector in the θ direction.

Again, we assume that the receptors are in equilibrium with the local ligand concentration, that binding and unbinding occur independently for each receptor, and that the cell is making a single observation at an instant in time. Thus, the probability that a particular binding configuration $(\mathcal{B}, \mathcal{U})$ will occur given $\boldsymbol{\mu}$, γ, and $\mathcal{R}$ is

$$P(\mathcal{B}, \mathcal{U}|\boldsymbol{\mu}, \gamma, \mathcal{R}) = C_{\mathcal{R}=\mathcal{B}\cup\mathcal{U}}(\mathcal{R}) \frac{\gamma^n}{(1+\gamma)^N} \prod_{\boldsymbol{r}\in\mathcal{B}} \frac{1 + \boldsymbol{r}^T\boldsymbol{\mu}}{1 + \rho\boldsymbol{r}^T\boldsymbol{\mu}} \prod_{\boldsymbol{r}\in\mathcal{U}} \frac{1}{1 + \rho\boldsymbol{r}^T\boldsymbol{\mu}}. \tag{2.58}$$

Applying the arguments made in the one-dimensional case to transition from equation 2.13 to equation 2.21, we obtain

$$P(\boldsymbol{\mu}|\mathcal{B}) \propto P(\boldsymbol{\mu}) \exp\left[(1-\rho)\boldsymbol{R_b}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T \left((1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]\right)\boldsymbol{\mu}\right], \tag{2.59}$$

where $\boldsymbol{R_b} = \sum_{r \in \mathcal{B}} \boldsymbol{r}$ and $\mathbf{S}_b = \sum_{r \in \mathcal{B}} \boldsymbol{r}\boldsymbol{r}^T$, and we assume that $(1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E[\boldsymbol{r}\boldsymbol{r}^T]$ is positive definite.

Given the posterior distribution $P(\boldsymbol{\mu}|\mathcal{B})$, the Bayes optimal policy maximizes the expected value of the utility function. We have:

$$\begin{aligned}
\pi_{bayes}(\mathcal{B}) &= \operatorname*{argmax}_{\theta} \int d\boldsymbol{\mu}\, \boldsymbol{u}_\theta . \boldsymbol{\mu}\, P(\boldsymbol{\mu}|\mathcal{B}) \\
&= \operatorname*{argmax}_{\theta} \boldsymbol{u}_\theta . E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|\mathcal{B})} \\
&= \operatorname{argmax} |E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|\mathcal{B})}| \cos(\theta - \angle E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|\mathcal{B})}) \\
&= \angle E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|BB)}\,, \qquad (2.60)
\end{aligned}$$

where $E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|\mathcal{B})}$ denotes the expectation value of $\boldsymbol{\mu}$ over the posterior and $\angle\boldsymbol{\mu}$ denotes the angle $\boldsymbol{\mu}$ forms with respect to the x-axis. In other words, the Bayes' optimal strategy states that the cell ought to select the direction of the mean a posteriori estimate of the gradient, $E[\boldsymbol{\mu}]_{P(\boldsymbol{\mu}|\mathcal{B})}$.

If we assume that $P(\boldsymbol{\mu})$ is radially symmetric around $\boldsymbol{\mu} = 0$ (i.e., the cell does not have any a priori bias toward any particular direction), the Bayes' optimal policy is to select the direction of $\boldsymbol{R_b}$. This is because the term $\exp[(1-\hat{\rho})\boldsymbol{R_b}.\boldsymbol{\mu}]$ in the posterior distribution increases most rapidly in the direction $\boldsymbol{\mu} = \boldsymbol{R_b}$, which skews probability mass to the side of the $\boldsymbol{\mu}$ prior distribution in the direction of $\boldsymbol{R_b}$.

More generally, we could consider prior distributions for $\boldsymbol{\mu}$ that are not radially symmetric around $\boldsymbol{\mu} = 0$. While there are many such possible distributions, we will limit ourselves to bivariate gaussian distributions. Assuming that $P(\boldsymbol{\mu})$ has a nonsingular covariance matrix $\mathbf{M}$, and mean $\boldsymbol{\mu}_0$, and that $(1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]$ is positive definite (which it almost surely is when the concentration is not too low), we have:

$$\begin{aligned}
P(\boldsymbol{\mu}|\mathcal{B}) &\propto \mathcal{N}(\boldsymbol{\mu};\, \boldsymbol{\mu}_0, \mathbf{M}) \exp\Bigg[(1-\rho)\boldsymbol{R_b}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T \\
&\qquad \left((1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]\right)\boldsymbol{\mu}\Bigg] \\
&\propto \mathcal{N}\left(\boldsymbol{\mu};\, \boldsymbol{\mu}_0', \mathbf{M}'\right), \qquad (2.61)
\end{aligned}$$

where

$$\boldsymbol{\mu}_0' = \left(1 + \left((1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]\right)\mathbf{M}\right)^{-1}(\mu_0 + \mathbf{M}(1-\rho)\boldsymbol{R_b}) \qquad (2.62)$$

$$\mathbf{M}' = \left(1 + \left((1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]\right)\mathbf{M}\right)^{-1}\mathbf{M}, \qquad (2.63)$$

and thus

$$\pi_{bayes}(\mathcal{B}) = \angle \left(1 + \left((1-\rho^2)\mathbf{S}_b - 2(N-n)\rho^2 E\left[\boldsymbol{r}\boldsymbol{r}^T\right]\right)\mathbf{M}\right)^{-1} \\ (\boldsymbol{\mu}_0 + (1-\rho)\mathbf{M}\boldsymbol{R}_b)\,. \tag{2.64}$$

Thus, the optimal policy combines both prior and sensory information in order to select a step direction. As the cell's prior certainty about the gradient direction (quantified by $|\mathbf{M}^{-1}\boldsymbol{\mu}_0|$) decreases, the relative importance of sensory information increases compared to prior information. Vice versa, when the cell's prior certainty is high, sensory input has reduced influence on the direction chosen.

2.5 Performance of the Two-Dimensional Bayes Optimal Policy. Given the gradient sensing strategy derived in the previous section, we would now like to estimate its performance over various environmental conditions for an entire ensemble of cells (described by the distribution of receptor arrangements, $P_r(\mathcal{R})$). In other words, we wish to determine how the expected utility, $E\left[G(\pi_{bayes}(\mathcal{B});\boldsymbol{\mu})\right]_{P(\mathcal{B}|\boldsymbol{\mu},\gamma,\mathcal{R}),P(\mathcal{R})}$ behaves as a function of $(\boldsymbol{\mu},\gamma)$. This involves two steps:

1. Obtaining an expression for the distribution of actions performed under the Bayes optimal policy for a given set of gradient conditions $(\boldsymbol{\mu},\gamma)$: $P(\theta = \pi_{bayes}(\mathcal{B})|\boldsymbol{\mu},\gamma)$
2. Finding the expectation of the utility function $E[G(\theta;\boldsymbol{\mu})]_{P(\theta|\boldsymbol{\mu},\gamma)}$ under this distribution.

We will focus on the unbiased prior case, in which the Bayes optimal policy is purely a function of $\boldsymbol{R}_b$. Thus, in order to obtain the distribution of actions under the policy, we proceed by estimating the distribution $P(\boldsymbol{R}_b|\boldsymbol{\mu},\gamma)$. As for the one-dimensional case, we approximate the distribution of $\boldsymbol{R}_b$ with a gaussian distribution:

$$P(\boldsymbol{R}_b|\gamma,\boldsymbol{\mu}) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{R}_b - E[\boldsymbol{R}_b]\right)^T \mathrm{Cov}(\boldsymbol{R}_b,\boldsymbol{R}_b)^{-1}\left(\boldsymbol{R}_b - E[\boldsymbol{R}_b]\right)\right], \tag{2.65}$$

where $E[\boldsymbol{R}_b]$ is the expectation value of $\boldsymbol{R}_b$ for given $\boldsymbol{\mu}$ and γ and $\mathrm{Cov}(\boldsymbol{R}_b,\boldsymbol{R}_b)$ is the covariance matrix of $\boldsymbol{R}_b$, again at fixed $\boldsymbol{\mu}$ and γ. This approximation can be justified through the central limit theorem, as $\boldsymbol{R}_b$ is a sum of many independent random variables.

Assuming $E[\boldsymbol{r}] = 0$, we have

$$E[\boldsymbol{R}_b] = NE[b\boldsymbol{r}] \tag{2.66}$$

$$= N\rho(1-\rho)\mathrm{Cov}(\boldsymbol{r},\boldsymbol{r})\boldsymbol{\mu} + O(\mu^2) \tag{2.67}$$

and

$$\mathrm{Cov}(\boldsymbol{R}_b, \boldsymbol{R}_b) = N\mathrm{Cov}(b\boldsymbol{r}, b\boldsymbol{r}) \tag{2.68}$$

$$= N\left(E\left[b\boldsymbol{r}\boldsymbol{r}^T\right] - E[b\boldsymbol{r}]\,E\left[b\boldsymbol{r}^T\right]\right) \tag{2.69}$$

$$= N\rho E\left[\boldsymbol{r}\boldsymbol{r}^T\right] + \rho(1-\rho)E\left[\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{r}^T\boldsymbol{\mu}\right] + O(\mu^2). \tag{2.70}$$

We now have the covariance and mean of $\boldsymbol{R}_b$ in terms of two functionals of the receptor distribution: $\mathrm{Cov}(\boldsymbol{r}, \boldsymbol{r})$ and $E\left[\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{r}^T\boldsymbol{\mu}\right]$. To proceed, we need to make assumptions about the form of $P_r(\boldsymbol{r})$. We do this for $P_r(\boldsymbol{r})$ elongated along a single axis (along the direction β) but centered on $\boldsymbol{r} = 0$. This includes as a special case a radially symmetric $P_r(\boldsymbol{r})$. This yields

$$\mathrm{Cov}(\boldsymbol{r}, \boldsymbol{r}) = E\left[\boldsymbol{r}\boldsymbol{r}^T\right] = \kappa^2\mathbf{R}_\beta \begin{bmatrix} \zeta^2 & 0 \\ 0 & \zeta^{-2} \end{bmatrix} \mathbf{R}_{-\beta} \tag{2.71}$$

$$E\left[\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{r}^T\right] = 0, \tag{2.72}$$

and hence the expectation and covariance of $\boldsymbol{R}_b$ are

$$E[\boldsymbol{R}_b] = \kappa^2 N\rho(1-\rho)\mathbf{R}_\beta \begin{bmatrix} \zeta^2 & 0 \\ 0 & \zeta^{-2} \end{bmatrix} \mathbf{R}_{-\beta}\boldsymbol{\mu}$$

$$\mathrm{Cov}(\boldsymbol{R}_b, \boldsymbol{R}_b) = \kappa^2 N\rho\mathbf{R}_\beta \begin{bmatrix} \zeta^2 & 0 \\ 0 & \zeta^{-2} \end{bmatrix} \mathbf{R}_{-\beta},$$

where ζ determines the strength of the elongation (without loss of generality and for later ease of presentation, we assume that $\zeta < 1$, which corresponds to elongation along the y-axis).

We now wish to estimate $P(\theta|\boldsymbol{\mu}, \gamma)$, the probability of the cell taking a step in the direction θ, given gradient conditions $\boldsymbol{\mu}$ and γ. Given $P(\boldsymbol{R}_b|\boldsymbol{\mu}, \gamma)$, we can obtain an expression for $P(\theta|\boldsymbol{\mu}, \gamma)$ by evaluating

$$P(\theta|\boldsymbol{\mu}, \gamma) = \int_{\{\boldsymbol{R}_b|\pi_{bayes}(\mathcal{B})=\theta\}} d^2\boldsymbol{R}_b\, P(\boldsymbol{R}_b|\boldsymbol{\mu}, \gamma), \tag{2.73}$$

where the integration is performed over the values of $\boldsymbol{R}_b$ for which the Bayes optimal policy elects to step in the direction θ. Making the approximation

$$P(\boldsymbol{R}_b = \boldsymbol{v}|\boldsymbol{\mu}, \gamma) \approx \frac{1}{2\pi\sqrt{|\mathrm{Cov}(\boldsymbol{R}_b, \boldsymbol{R}_b)|}}$$

$$\times \exp\left[-\frac{1}{2}(\boldsymbol{v} - E[\boldsymbol{R}_b])\,\mathrm{Cov}(\boldsymbol{R}_b, \boldsymbol{R}_b)^{-1}(\boldsymbol{v} - E[\boldsymbol{R}_b])\right], \tag{2.74}$$

where $|\mathrm{Cov}(\boldsymbol{R_b}, \boldsymbol{R_b})|$ is the determinant of $\mathrm{Cov}(\boldsymbol{R_b}, \boldsymbol{R_b})$ and substituting 2.74 into 2.73, we obtain:

$$P(\theta|\boldsymbol{\mu}, \gamma) = \frac{1}{2\pi\sqrt{|\mathrm{Cov}(\boldsymbol{R_b}, \boldsymbol{R_b})|}} \int_0^\infty dr\, r \times \exp\left[-\frac{1}{2}\left(r\boldsymbol{u}_\theta - E[\boldsymbol{R_b}]\right)^T \mathrm{Cov}(\boldsymbol{R_b}, \boldsymbol{R_b})^{-1}\left(r\boldsymbol{u}_\theta - E[\boldsymbol{R_b}]\right)\right]. \quad (2.75)$$

To proceed, we incorporate the expressions for $E[\boldsymbol{R_b}]$ and $\mathrm{Cov}(\boldsymbol{R_b}, \boldsymbol{R_b})$ obtained above. Defining

$$\begin{aligned} A &= \boldsymbol{u_\theta}^T \mathrm{Cov}\left(\boldsymbol{R_b}, \boldsymbol{R_b}\right)^{-1} \boldsymbol{u_\theta} \\ &= \kappa^{-2} N^{-1} \rho^{-1} \boldsymbol{u_\theta}^T \mathbf{R}_\beta \mathrm{diag}\left(\zeta^{-2}, \zeta^2\right) \mathbf{R}_{-\beta} \boldsymbol{u_\theta}, \quad (2.76) \\ B &= E[\boldsymbol{R_b}]^T \mathrm{Cov}\left(\boldsymbol{R_b}, \boldsymbol{R_b}\right)^{-1} \boldsymbol{u_\theta} \\ &= (1-\rho)\, \boldsymbol{\mu}^T \boldsymbol{u_\theta}, \quad (2.77) \end{aligned}$$

and

$$\begin{aligned} C &= E[\boldsymbol{R_b}]^T \mathrm{Cov}\left(\boldsymbol{R_b}, \boldsymbol{R_b}\right)^{-1} E[\boldsymbol{R_b}] \\ &= \kappa^2 N \rho\, (1-\rho)^2\, \boldsymbol{\mu}^T \mathbf{R}_\beta \mathrm{diag}\left(\zeta^2, \zeta^{-2}\right) \mathbf{R}_{-\beta} \boldsymbol{\mu}, \quad (2.78) \end{aligned}$$

we can rewrite equation 2.75 in the form

$$\begin{aligned} &\int_0^\infty dr\; r\; \exp\left(-\frac{1}{2}\left(Ar^2 - 2Br + C\right)\right) \\ &\quad = \frac{1}{A} \exp\left(-\frac{1}{2}C\right)\left(1 + \sqrt{\pi}\, Q \exp Q^2 \left(1 + \mathrm{erf}\left(Q\right)\right)\right), \quad (2.79) \end{aligned}$$

where $Q = \frac{B}{\sqrt{2A}}$.

Without loss of generality, we can assume that $\beta = 0$, since it is only the relative angle, $\phi = \angle\boldsymbol{\mu} - \beta$, between the true direction of the gradient and the axis along which the cell is elongated that should enter into the final expression. This simplifies our expressions to

$$\begin{aligned} A &= \kappa^{-2} N^{-1} \rho^{-1} u_\theta^T \mathrm{diag}\left(\zeta^{-2}, \zeta^2\right) \boldsymbol{u_\theta} \\ &= \kappa^{-2} N^{-1} \rho^{-1} \chi^2\left(\theta, \zeta\right), \quad (2.80) \\ C &= \kappa^{-2} N \rho\, (1-\rho)^2\, \boldsymbol{\mu}^{\mathrm{T}} \mathrm{diag}\left(\zeta^2, \zeta^{-2}\right) \boldsymbol{\mu} \\ &= d^2 \chi^2\left(\phi, \zeta^{-1}\right), \quad (2.81) \end{aligned}$$

and

$$Q = \kappa \sqrt{\frac{N}{2}} \sqrt{\rho}\,(1-\rho)\,|\boldsymbol{\mu}|\,\frac{\boldsymbol{u}_{\angle\boldsymbol{\mu}}^{T}\boldsymbol{u}_{\boldsymbol{\theta}}}{\sqrt{\boldsymbol{u}_{\boldsymbol{\theta}}{}^{T}\,\mathrm{diag}\,(\zeta^{-2},\zeta^{2})\,\boldsymbol{u}_{\boldsymbol{\theta}}}}$$
$$= d'\frac{\cos(\theta-\phi)}{\chi(\theta,\zeta)}, \tag{2.82}$$

where we have defined

$$\chi(\theta,\zeta) = \sqrt{\boldsymbol{u}_{\boldsymbol{\theta}}{}^{T}\,\mathrm{diag}\,(\zeta^{-2},\zeta^{2})\,\boldsymbol{u}_{\boldsymbol{\theta}}} \tag{2.83}$$

and a signal-to-noise ratio

$$d' = \kappa\sqrt{\frac{N}{2}}\,|\boldsymbol{\mu}|\,\sqrt{\rho}(1-\rho)$$
$$= \kappa\sqrt{\frac{N}{2}}\mathrm{SNR}_{\boldsymbol{B}}, \tag{2.84}$$

where $\mathrm{SNR}_{\boldsymbol{B}}$ is the signal-to-noise ratio defined in equation 2.26 in section 2.1.

Substituting everything into equation 2.75, we have:

$$P(\theta|\boldsymbol{\mu},\gamma) = \frac{1}{2\pi}\frac{\exp\left(-d'^{2}\chi^{2}(\alpha,\zeta^{-1})/2\right)}{\chi^{2}(\theta,\zeta)}$$
$$\times\left(1+\sqrt{\pi}d'\frac{\cos(\theta-\phi)}{\chi(\theta,\zeta)}\exp\left(d'^{2}\frac{\cos^{2}(\theta-\phi)}{\chi^{2}(\theta,\zeta)}\right)\right.$$
$$\left.\times\left(1+\mathrm{erf}\left(d'\frac{\cos(\theta-\phi)}{\chi(\theta,\zeta)}\right)\right)\right)$$
$$\approx \frac{1}{2\sqrt{\pi}}d'\frac{\cos(\theta-\phi)}{\chi^{3}(\theta,\zeta)} \tag{2.85}$$

to first order in the signal-to-noise ratio d'.

It is then straightforward to estimate the performance of the optimal strategy in terms of the expected value of the utility function:

$$E[G(\theta;\boldsymbol{\mu})] = |\boldsymbol{\mu}|\int d\theta\,\cos(\theta-\angle\boldsymbol{\mu})P(\theta|\boldsymbol{\mu},\gamma). \tag{2.86}$$

After incorporating equation 2.85, this integral simplifies to

$$E[G(\theta;\boldsymbol{\mu},\gamma)] = |\boldsymbol{\mu}|\frac{1}{2\sqrt{\pi}}d'\int_0^{2\pi} d\theta \frac{\cos^2(\theta-\phi)}{\chi^3(\theta,\zeta)}.$$

For $\zeta = 1$, this can be evaluated immediately to give

$$= |\boldsymbol{\mu}|\frac{\sqrt{\pi}}{2}d', \tag{2.87}$$

while for $\zeta < 1$, we have

$$E[G(\theta;\boldsymbol{\mu},\gamma)] = |\boldsymbol{\mu}|\frac{1}{4\sqrt{\pi}}d' \times \int_0^{2\pi} d\theta \frac{1+\cos 2\theta\cos 2\phi+\sin 2\phi}{\chi^3(\theta,\zeta)} \tag{2.88}$$

Symmetry arguments show that the $\sin 2\phi$ term integrates to zero, and the entire expression reduces to

$$E[G(\theta;\boldsymbol{\mu},\gamma)] = |\boldsymbol{\mu}|\frac{1}{\sqrt{\pi}}d' \times (I_0(\zeta)+I_1(\zeta)\cos 2\phi) \tag{2.89}$$

where the integrals

$$I_0(\zeta) = \int_0^{\pi/2} d\theta \frac{1}{\chi^3(\theta,\zeta)} = \frac{1}{\zeta}E\left(\sqrt{1-\zeta^4}\right) \tag{2.90}$$

and

$$I_1(\zeta) = \int_0^{\pi/2} d\theta \frac{\cos 2\theta}{\chi^3(\theta,\zeta)} = 2\frac{\zeta^3}{1-\zeta^4}K\left(\sqrt{1-\zeta^4}\right) + \frac{1}{\zeta}E\left(\sqrt{1-\zeta^4}\right) \tag{2.91}$$

can be evaluated in terms of the complete elliptic integrals $K(v) = \int_0^{\pi/2} d\theta(1-v^2\sin^2\theta)^{-1/2}$ and $E(v) = \int_0^{\pi/2} d\theta(1-v^2\sin^2\theta)^{-1/2}$ (note, however, that this form is correct only for $\zeta < 1$; results for $\zeta > 1$ can be obtained by rotating the coordinate systems by $\pi/2$).

There are several noteworthy features of equations 2.87 and 2.89. First, note that performance is proportional to $\sqrt{\rho}(1-\rho)$. This is exactly the result we obtain for the analogous one-dimensional situation—that is, when solely bound receptors signal. Hence, our predictions concerning the shape of the performance curve (with respect to concentration) are the same for both the one-dimensional and two-dimensional situations. Similarly, we observe the same dependence on receptor number and variance in the receptor distribution, κ^2, with performance proportional to $\kappa\sqrt{N}$. Thus, the

basic conclusions from the one-dimensional case carry over to the two-dimensional case.

3 Integrating Information over Time

So far we have considered the cell's ability to detect a gradient based on only the states of its receptors at a single instant in time. However, in order to interact optimally with its environment, the cell must integrate information from its receptors over time. In this situation, the dynamics of the environment play an important role in determining the timescale on which information remains relevant. At a concrete level, this modifies the Bayesian analysis we have presented so far, as the cell's prior must now also describe its knowledge of the dynamics of its environment. Under the assumptions that

- the cell makes inferences at a series of discrete (dimensionless) times $t = 1, 2, \ldots$,
- these times are sufficiently separated that correlations between receptor binding states at adjacent times are negligible, and
- that the state of the environment at time $t + 1$ can depend only on the state of the environment at time t,

we can formulate the problem as estimating the state of a hidden Markov model, to which we apply recursive Bayesian filtering.

Under this formulation, we are trying to estimate the gradient $\boldsymbol{\mu}_t$ at time t, assuming known dynamics $P(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1})$ based on observations $\boldsymbol{\mathcal{B}}_t = \mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_t$ made up to and including the current time. The problem is to obtain the distribution $P(\boldsymbol{\mu}_t|\boldsymbol{\mathcal{B}}_t)$ of possible states given the observations. It can be shown that this can be obtained iteratively through

$$P(\boldsymbol{\mu}_t|\boldsymbol{\mathcal{B}}_t) = \frac{P(\mathcal{B}_t|\boldsymbol{\mu}_t)P(\boldsymbol{\mu}_t|\boldsymbol{\mathcal{B}}_{t-1})}{P(\mathcal{B}_t|\boldsymbol{\mathcal{B}}_{t-1})} \tag{3.1}$$

and

$$P(\boldsymbol{\mu}_t|\boldsymbol{\mathcal{B}}_{t-1}) = \int d^2\boldsymbol{\mu}_{t-1} P(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1})P(\boldsymbol{\mu}_{t-1}|\boldsymbol{\mathcal{B}}_{t-1}). \tag{3.2}$$

Equation 3.1 is simply an application of equation 2.59, with $P(\boldsymbol{\mu}_t|\boldsymbol{\mathcal{B}}_{t-1})$ acting as the prior distribution. Assuming that $P(\boldsymbol{\mu}_0)$ is normally distributed and that the gradient steepness vector drifts according to a diffusion process, we can express equations 3.1 and 3.2 as transformations of multivariate normal distributions:

$$\mathcal{N}\left(\boldsymbol{\mu}_t; \boldsymbol{\mu}_0^{(t)}, \mathbf{M}^{(t)}\right) \to \mathcal{N}\left(\boldsymbol{\mu}_t; \boldsymbol{\mu}_0^{(t)'}, \mathbf{M}^{(t)'}\right) \tag{3.3}$$

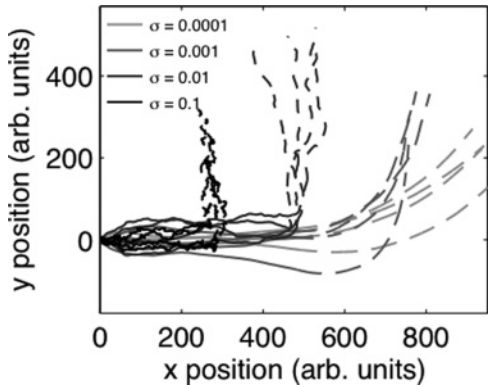

Figure 9: Expected rate of environmental change determines sensitivity to fluctuations and interpolates between unpolarized and polarized behavior. Each trajectory was generated by repeatedly generating binding distributions $\mathcal{B}$, according to the specified gradient conditions, applying equations 3.1 and 3.2, then taking a unit step in the direction mandated by equation 2.60. For each trajectory, the gradient is pointing to the right for the first 500 time steps (shown as solid lines), and pointing up for the second 500 time steps (shown as dotted lines). Depending on the expected rate σ^2 of drift of the gradient steepness vector, the cell either responds slowly to changes in the gradient direction (as illustrated by the $\sigma = 0.0001$ case) or responds immediately to changes in gradient direction at the potential cost of being overly sensitive to stochastic fluctuations (e.g., $\sigma = 0.1$). In each case, $|\boldsymbol{\mu}| = 0.1$ and $\rho = 0.5$.

$$\mathcal{N}\left(\boldsymbol{\mu}_{t-1}; \boldsymbol{\mu}_0^{(t-1)}, \mathbf{M}^{(t-1)}\right) \to \mathcal{N}\left(\boldsymbol{\mu}_t; \boldsymbol{\mu}_0^{(t-1)}, \mathbf{M}^{(t-1)} + \sigma^2\right), \tag{3.4}$$

where $\boldsymbol{\mu}_0^{(t)'}$ and $\mathbf{M}^{(t)'}$ are as defined in equations 2.62 and 2.63 and σ^2 defines the rate at which the cell expects the gradient steepness vector to drift.

Figure 9 illustrates the influence of the parameter σ^2 on trajectory of a cell exposed to a sudden rapid change in gradient direction. Cells that expect a dynamic, rapidly changing environment (σ large) display an immediate change in their direction of travel, while cells expecting a slowly changing environment tend to display smoother "turning" behavior. These behaviors are reminiscent of the behavior of cells with different levels of polarization: unpolarized cells tend to display random walk behavior, responding rapidly to any changes in the gradient direction, while polarized cells tend to change direction by executing smooth turns.

We should note that the model of environmental dynamics we have used in this section is highly simplified: our intent was to demonstrate that chemotactic behavior reminiscent of that of polarized and unpolarized cells can emerge naturally from the Bayesian framework. The model we have focused on here assumes that the direction and strength of the gradient change

independent of any action the cell might take and that the background concentration is constant. While this situation may be approximately the case when, for example, the cell is attempting to follow a moving concentration source, a more sophisticated environmental model would need to include predicted changes in the background concentration due to the cell's motion. Strong et al. (1998) demonstrate more complex environmental models of this sort in the context of bacterial chemotaxis.

4 Discussion

Other recent work has also focused on the role of noise in eukaryotic chemotaxis and gradient sensing (e.g. Ueda & Shibata, 2007; van Haastert & Postma, 2007; Andrews & Iglesias, 2007; Endres & Wingreen, 2008; Fuller et al., 2010). Our approach differs from these studies in that we focus on understanding how to optimally combine information from spatially distributed receptors and how various assumptions about receptor signaling (i.e., whether bound or unbound receptors signal) influence gradient sensing ability. However, we should note that there are several important phenomena that our model does not incorporate. These include, for example, receptor trafficking (Aquino & Endres, 2010), receptor clustering (Bray, Levin, & Morton-Firth, 1998) and intrinsic noise in receptor signaling (Ueda & Shibata, 2007).

An important issue with the results presented here (and by other authors: Andrews & Iglesias, 2007; Fuller et al., 2010) is that we ignore temporal (Mortimer, Dayan, Burrage, & Goodhill, 2009; Endres & Wingreen, 2009) and spatiotemporal correlations (Berg & Purcell, 1977; Bialek & Setayeshgar, 2005), in particular, their dependence on concentration. Our model implicitly assumes that the rate at which the cell's sensory apparatus makes independent measurements of the gradient has no dependence on concentration. However, studies estimating the accuracy with which cells can measure absolute concentration find that the time over which the binding state of receptors remains correlated is inversely proportional to the background concentration, which suggests that independent measurements can be made faster as concentration increases (Berg & Purcell, 1977; Bialek & Setayeshgar, 2005; Endres & Wingreen, 2009). This would mean that our model underestimates performance at high concentrations. Indeed, we have shown that when the cell has access to the full trajectory through time of its receptors, its optimal gradient sensing performance is monotonically increasing with concentration (Mortimer, Dayan, et al., 2009). One effect that may provide a lower bound on the correlation time of gradient measurements is the rate at which receptors are mixed on the surface of the cell. If only bound receptors signal, fluctuations in the receptor distribution bias the gradient measurement. At high concentrations, these dominate gradient-induced asymmetries in receptor-generated signals, and thus the correlation time would be dominated by the rate of receptor mixing.

What does the two-dimensional Bayesian analysis we have presented here suggest for biological chemotactic systems? Observations from *D. discoideum* and neutrophils have led to the broadly accepted view that three related phenomena underlie eukaryotic chemotaxis (van Haastert & Devreotes, 2004; Willard & Devreotes, 2006; Mortimer et al., 2008):

- External signals are transduced through a directional sensing mechanism, by which a shallow external gradient of guidance molecules drives the production of a significantly steeper intracellular gradient of activity. This is achieved through two submodules: an amplification module, which magnifies weak intracellular asymmetries through positive feedback, and an adaptation module, which ensures that gradient sensing works effectively over a wide range of background concentrations (Levchenko & Iglesias, 2002).
- This intracellular activity in turn regulates cytoskeletal dynamics and cell adhesion, leading to motility.
- Motility and directional sensing occur in the context of cell or growth cone polarization, or the persistent arrangement of key signaling molecules into distinct "front" and "rear" regions.

We suggest, in direct analogy to the decision theory framework, that directional sensing might be thought of as decision making. Under this interpretation, the amplification and adaptation submodules act to make an unambiguous estimate of the gradient direction, through a winner-takes-all mechanism promoting pseudopod extension only at points on the periphery that exhibit sufficiently high levels of receptor-mediated signaling. As amplification processes are generally highly nonlinear but subject to saturation effects, they would act most effectively at a specific input level—a set point. Adaptation achieved through "local excitation, global inhibition" (Parent & Devreotes, 1999; Levchenko & Iglesias, 2002) would maintain global signaling levels near this set point, thus maximizing the effectiveness of the amplification module. Under this conceptual framework, amplification of a weak asymmetry is considered a decision-making process, leading to the cell or growth cone extending a process in a specific direction to enact a turn.

Thinking of amplification as decision making also allows an intuitive decision-theoretic interpretation to cell polarization. Polarized cells tend to be significantly more sensitive to chemotactic stimulation at their leading edge. An interpretation of this is that in polarized cells, the cell already "knows" roughly where the gradient is pointing, and thus its response tends to be a local change in this direction, rather than rapid turning. From the perspective of statistical decision theory, we can identify the polarization of a cell with an asymmetric prior distribution for the gradient direction. This is also consistent with experiments in which *D. discoideum* cells were transiently stimulated by growing them in a medium containing caged cAMP, then using a short laser pulse to uncage the cAMP (Samadani, Mettetal, & van Oudenaarden, 2006). In this study, *Dictyostelium* cells were observed

to show more or less 'bias' towards some apparently random prior direction. Extending this experimental paradigm would provide one test of our model: if the cells were exposed to a gradient pointing in a single direction for a long period, we would expect any underlying bias in the response to a point stimulation by optical uncaging to be aligned with the previously imposed gradient. We would also predict the strength of the underlying biases to be related to the time for which the cells were exposed to the external gradient.

Indeed, Andrews and Iglesias (2007) recently demonstrated through an approach based on rate-distortion theory that a near-optimal rate-distortion trade-off can be achieved by such a strategy, given sufficient downstream amplification. A distinction between our approach and that of Andrews and Iglesias (2007) is that we directly characterize how a cell can make best use of all of the information it receives through its receptors in order to make its decision. In contrast, the definition of optimality taken by Andrews and Iglesias (2007) cannot distinguish between a cell that extracts all possible information from its receptors and one that, for example, ignores the signals from a uniform subset of its receptors: both strategies will yield the same rate-distortion curve, but the second is certainly not making optimal use of the information from all of its receptors.

What might be the biological substrate underlying this memory? In polarized *Dictyostelium*, G-proteins are distributed in a graded way, so that more are found at the leading than trailing edge (Jin, Zhang, Long, Parent, & Devreotes, 2000). Under this scheme, the concentration of G-protein at a particular point on the periphery is proportional to the prior probability that the gradient is increasing in that direction. A simple model in which bound receptors noncooperatively recruit and activate G-proteins then immediately results in downstream activity proportional to the posterior distribution: the product of likelihood and prior. An approximation to the statistically optimal decision would then follow through downstream amplification.

In conclusion, we have extended the one-dimensional model presented first in Mortimer, Feldner, et al. (2009) in several ways. We have shown that based on this model, there is a limit to the distance over which a cell can be guided by a single gradient. We have extended our model to treat the case of guidance by multiple correlated gradients and have demonstrated that optimally integrating information from multiple gradients requires a "weighting" step in order to account for variability in source reliability. Finally, we have extended our model to two dimensions, demonstrating both that the key features observed in the one-dimensional model are unchanged and that we can reproduce behavior observed in several cell types. Specifically, we have demonstrated a clear connection between prior information about gradient direction and cell polarization. The optimal strategy for a cell or growth cone with no prior information about the gradient direction leads to behavior reminiscent of that of an unpolarized cell, while the optimal strategy assuming strong prior information about

the gradient direction leads to behavior consistent with that of polarized cells.

Acknowledgments

This work was supported by an Australian Postgraduate Award (to D.M.), an Australian Research Council Federation Fellowship (to K.B.), the Gatsby Charitable Foundation (to P.D.), Program Grant RPG0029/2008-C from the Human Frontier Science Program, Australian Research Council Discovery Grant DP0666126 and Australian National Health and Medical Research Council Project Grant 456003. The authors would like to thank Clement Bonini for helpful comments on the manuscript.

References

Abramowitz, M., & Stegun, I. A. (Eds.). (1970). *Handbook of mathematical functions*. New York: Dover.

Aeschlimann, M., & Tettoni, L. (2001). Biophysical model of axonal pathfinding. *Neurocomput., 38*, 87–92.

Andrews, B. W., & Iglesias, P. A. (2007). An information-theoretic characterization of the optimal gradient sensing response of cells. *PLoS Comput. Biol., 3*, 1490–1497.

Aquino, G., & Endres, R. G. (2010). Increased accuracy of ligand sensing by receptor internalization. *Phys. Rev. E., 81*, 021909.

Arrieumerlou, C. & Meyer, T. (2005). A local coupling model and compass parameter for eukaryotic chemotaxis. *Dev. Cell, 8*, 215–227.

Berg, H. C., & Purcell, E. M. (1977). Physics of chemoreception. *Biophys. J., 20*, 193–219.

Bialek, W., & Setayeshgar, S. (2005). Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. USA, 102*, 10040–10045.

Bray, D., Levin, M. D., & Morton-Firth, C. J. (1998). Receptor clustering as a cellular mechanism to control sensitivity. *Nature, 393*, 85–88.

Charron, F., Stein, E., Jeong, J., McMahon, A. P., & Tessier-Lavigne, M. (2003). The morphogen Sonic Hedgehog is an axonal chemoattractant that collaborates with Netrin-1 in midline axon guidance. *Cell, 113*, 11–23.

DeLisi, C., & Marchetti, F. (1982). A theory of measurement error and its implications for spatial and temporal gradient sensing during chemotaxis. *Cell Biophys., 4* 211–229.

Downey, G. P. (1994). Mechanisms of leukocyte motility and chemotaxis. *Curr. Opin. Immunol., 6*, 113–124.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.

Endres, R. G., & Wingreen, N. S. (2008). Accuracy of direct gradient sensing by single cells. *Proc. Natl. Acad. Sci. USA, 105*, 15749–15754.

Endres, R. G., & Wingreen, N. S. (2009). Maximum likelihood and the single receptor. *Phys. Rev. Lett., 103*, 158101.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Fisher, P. R., Merkl, R., & Gerisch, G. (1989). Quantitative analysis of cell motility and chemotaxis in *dictyostelium discoideum* by using an image processing system and a novel chemotaxis chamber providing stationary chemical gradients. *J. Cell Biol., 108*, 973–984.

Fuller, D., Chen, W., Adler, M., Groisman, A., Levine, H., Rappel, W.-J., et al. (2010). External and internal constraints on eukaryotic chemotaxis. *Proc. Natl. Acad. Sci. USA, 107*, 9656–9659.

Goodhill, G. J. (1997). Diffusion in axon guidance. *Eur. J. Neurosci., 9*, 1414–1421.

Goodhill, G. J., & Baier, H. (1998). Axon guidance: Stretching gradients to the limit. *Neur. Comput., 10,* 521–527.

Goodhill, G., & Urbach, J. (1999). Theoretical analysis of gradient detection by growth cones. *J. Neurobiol., 41*, 230–241.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Hoboken, NJ: Wiley.

Hao, J. C., Yu, T. W., Fujisawa, K., Culotti, J. G., Gengyo-Ando, K., Mittani, S., et al. (2001). *C. elegans* slit acts in midline, dorsal-ventral, and anterior-posterior guidance via the SAX-3/Robo receptor. *Neuron, 32*, 25–38.

Heit, B., Tavener, S., Raharjo, E., & Kubes, P. (2002). An intracellular signaling hierarchy determines direction of migration in opposing chemotactic gradients. *J. Cell Biol., 159*, 91–102.

Hentschel, H. G., & van Ooyen, A. (1999). Models of axon guidance and bundling during development. *Proc. Biol. Sci., 266*, 2231–2238.

Jin, T., Zhang, N., Long, Y., Parent, C. A., & Devreotes, P. N. (2000). Localization of the G protein betagamma complex in living cells during chemotaxis. *Science, 287*, 1034–1036.

Kimmel, J. M., Salter, R. M., & Thomas, P. J. (2007). An information theoretic framework for eukaryotic gradient sensing. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems, 19* (pp. 705–712). Cambridge, MA: MIT Press.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci., 27*, 712–719.

Kohout, T. A., & Lefkowitz, R. J. (2003). Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization. *Mol. Pharmacol., 63*, 9–18.

Kording, K. (2007). Decision theory: What "should" the nervous system do? *Science, 318*, 606–610.

Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244–247.

Lauffenburger, D. A. (1982). Influence of external concentration fluctuations on leukocyte chemotactic orientation. *Cell Biophys., 4*, 177–209.

Lauffenburger, D. A., & Linderman, J. J. (1996). *Receptors: Models for binding, trafficking and signaling*. New York: Oxford University Press.

Levchenko, A., & Iglesias, P. A. (2002). Models of eukaryotic gradient sensing: Application to chemotaxis of amoebae and neutrophils. *Biophys. J., 82*, 50–63.

Levine, H., Kessler, D. A., & Rappel, W. J. (2004). Directional sensing in eukaryotic chemotaxis: A balanced inactivation model. *Proc. Natl. Acad. Sci. USA, 101*, 8951–8956.

Lowery, L. A., & van Vactor, D. (2009). The trip of the tip: Understanding the growth cone machinery. *Nat. Rev. Mol. Cell. Biol., 10*, 332–343.

McLaughlin, T., & O'Leary, D. D. M. (2005). Molecular gradients and development of retinotopic maps. *Annu. Rev. Neurosci., 28*, 327–355.

Mortimer, D., Dayan, P., Burrage, K., & Goodhill, G. J. (2009). Optimizing chemotaxis by measuring unbound-bound transitions. *Physica D, 239*, 477–484.

Mortimer, D., Feldner, J., Vaughan, T., Vetter, I., Pujic, Z., Rosoff, W. J., et al. (2009). A Bayesian model predicts the response of axons to molecular gradients. *Proc. Natl. Acad. Sci. USA, 106*, 10296–10301.

Mortimer, D., Fothergill, T., Pujic, Z., Richards, L. J., & Goodhill, G. J. (2008). Growth cone chemotaxis. *Trends Neurosci., 31*, 90–98.

Naoki, H., Sakumura, Y., & Ishii, S. (2008). Stochastic control of spontaneous signal generation for gradient sensing in chemotaxis. *J. Theor. Biol., 255*, 259–266.

Narang, A. (2006). Spontaneous polarization in eukaryotic gradient sensing: A mathematical model based on mutual inhibition of frontness and backness pathways. *J. Theor. Biol., 240*, 538–553.

Onsum, M., & Rao, C. V. (2007). A mathematical model for neutrophil gradient sensing and polarization. *PLoS Comput. Biol., 3*, e36.

Parent, C. A., & Devreotes, P. N. (1999). A cell's sense of direction. *Science, 284*, 765–770.

Sakumura, Y., Tsukada, Y., Yamamoto, N., & Ishii, S. (2005). A molecular model for axon guidance based on cross talk between rho GTPases. *Biophys. J., 89*, 812–822.

Samadani, A., Mettetal, J., & van Oudenaarden, A. (2006). Cellular asymmetry and individuality in directional sensing. *Proc. Natl. Acad. Sci. USA, 103*, 11549–11554.

Skupsky, R., Losert, W., & Nossal, R. J. (2005). Distinguishing modes of eukaryotic gradient sensing. *Biophys. J., 89*, 2806–2823.

Skupsky, R., McCann, C., Nossal, R., & Losert, W. (2007). Bias in the gradient-sensing response of chemotactic cells. *J. Theor. Biol., 247*, 242–258.

Song, H., & Poo, M.-M. (2001). The cell biology of neuronal navigation. *Nat. Cell. Biol., 3*, E81–E88.

Strong, S. P., Freedman, B., Bialek, W., & Koberle, R. (1998). Adaptation and optimal chemotactic strategy for *E. coli*. *Phys. Rev. E., 57*, 4604–4617.

Tani, T., Miyamoto, Y., Fujimori, K. E., Taguchi, T., Yanagida, T., Sako, Y., et al. (2005). Trafficking of a ligand-receptor complex on the growth cones as an essential step for the uptake of nerve growth factor at the distal end of the axon: A single-molecule analysis. *J. Neurosci., 25*, 2181–2191.

Tranquillo, R. T., & Lauffenburger, D. A. (1986). Consequences of chemosensory phenomena for leukocyte orientation. *Cell Biophys., 8*, 1–46.

Tranquillo, R. T., & Lauffenburger, D. A. (1987). Stochastic-model of leukocyte chemosensory movement. *J. Math. Biol., 25*, 229–262.

Ueda, M., Sako, Y., Tanaka, T., Devreotes, P., & Yanagida, T. (2001). Single-molecule analysis of chemotactic signaling in *Dictyostelium* cells. *Science, 294*, 864–867.

Ueda, M., & Shibata, T. (2007). Stochastic signal processing and transduction in chemotactic response of eukaryotic cells. *Biophys. J., 93*, 11–20.

van Haastert, P. J. M., & Devreotes, P. N. (2004). Chemotaxis: Signalling the way forward. *Nat. Rev. Mol. Cell Biol., 5*, 626–634.

van Haastert, P. J. M., & Postma, M. (2007). Biased random walk by stochastic fluctuations of chemoattractant-receptor interactions at the lower limit of detection. *Biophys. J., 93*, 1787–1796.

Willard, S. S., & Devreotes, P. N. (2006). Signaling pathways mediating chemotaxis in the social amoeba, *Dictyostelium discoideum*. *Eur. J. Cell Biol., 85*, 897–904.

Zigmond, S. H. (1981). Consequences of chemotactic peptide receptor modulation for leukocyte orientation. *J. Cell Biol., 88*, 644–647.

Received June 21, 2010; accepted July 2, 2010.